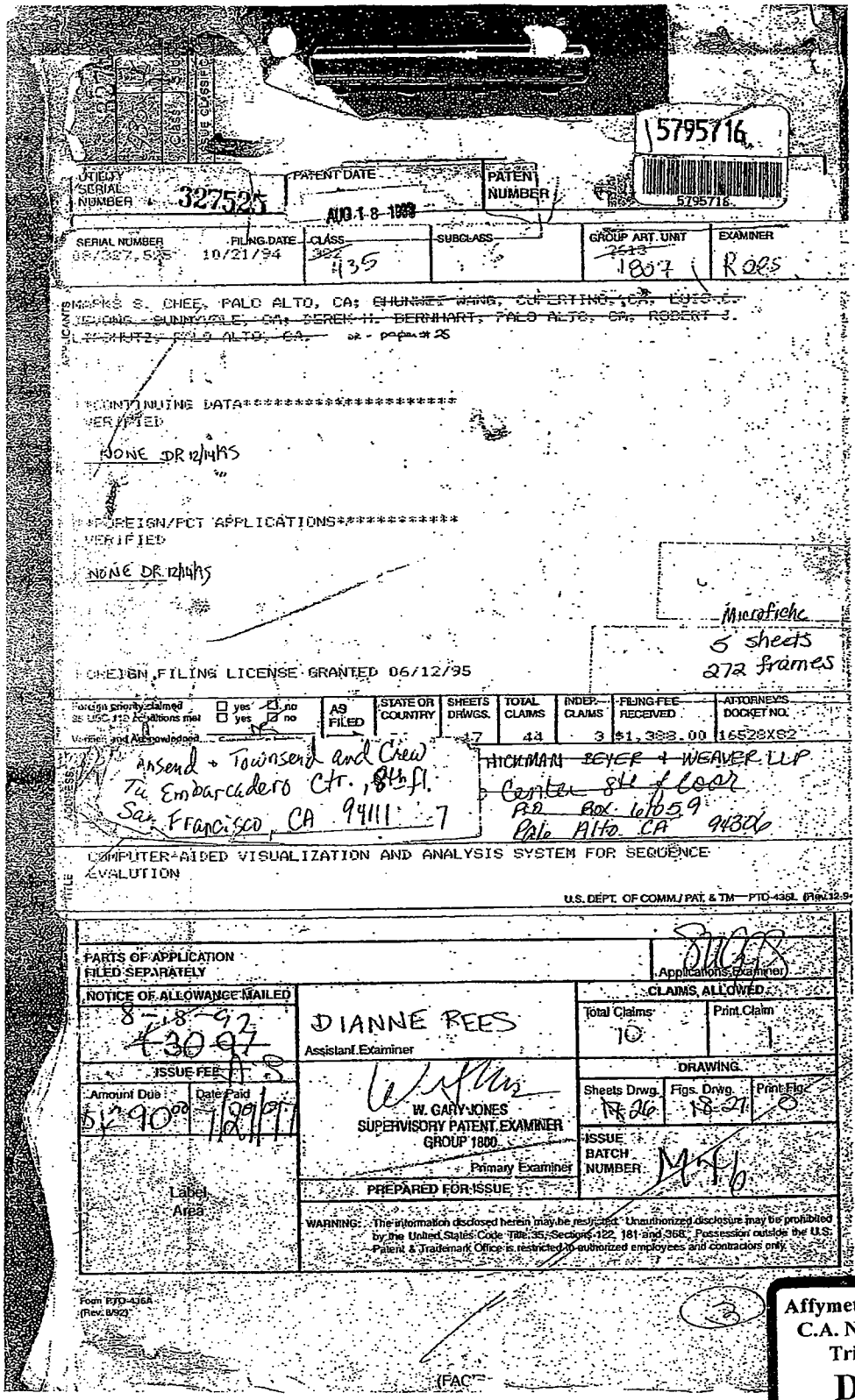

5795716

UTILITY SERIAL NUMBER 327525 | PATENT DATE AUG 18 1998 | PATENT NUMBER 5795716

SERIAL NUMBER	FILING DATE	CLASS	SUBCLASS	GROUP ART UNIT	EXAMINER
08/327,525	10/21/94	395 435		2613 1807	ROBS

APPLICANTS: MARK S. CHEE, PALO ALTO, CA; ~~CHUNNIS WANG, CUPERTINO, CA; LOIS L.~~ ~~HSIUNG, SUNNYVALE, CA; DEREK H. BERNHART, PALO ALTO, CA; ROBERT J.~~ ~~LEVINTHAL, PALO ALTO, CA.~~ ** - page # 25

CONTINUING DATA*****
VERIFIED

NONE DR 12/14/95

FOREIGN/PCT APPLICATIONS*****
VERIFIED

NONE DR 12/14/95

Microfiche
5 sheets
272 frames

FOREIGN FILING LICENSE GRANTED 06/12/95

Foreign priority claimed 35 USC 119 conditions met	Verified and Acknowledged	ASI FILED	STATE OR COUNTRY	SHEETS DRWGS.	TOTAL CLAIMS	INDEP. CLAIMS	FILING FEE RECEIVED	ATTORNEYS DOCKET NO.
☐ yes ☐ no				17	44	3	$1,388.00	16528XS2

ADDRESS: ~~Heisend + Townsend and Crew~~ ~~Two Embarcadero Ctr., 8th fl.~~ ~~San Francisco, CA 94111-3~~ ~~HICKMAN BEYER & WEAVER LLP~~ ~~Center 8th floor~~ ~~PO Box 16159~~ ~~Palo Alto CA 94306~~

TITLE: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUTION

U.S. DEPT. OF COMM./ PAT. & TM—PTO-436L (Rev.12-9

PARTS OF APPLICATION FILED SEPARATELY | Applications Examiner

NOTICE OF ALLOWANCE MAILED: 8-18-98 ~~$300~~

ISSUE FEE: Amount Due ~~$290.00~~ | Date Paid 12/9/98

Label Area

DIANNE REES
Assistant Examiner

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
Primary Examiner

PREPARED FOR ISSUE

CLAIMS ALLOWED: Total Claims 10 | Print Claim 1

DRAWING: Sheets Drwg. ~~26~~ | Figs. Drwg. ~~18-21~~ | Print Fig. 8

ISSUE BATCH NUMBER: M46

WARNING: The information disclosed herein may be restricted. Unauthorized disclosure may be prohibited by the United States Code Title 35, Sections 122, 181 and 368. Possession outside the U.S. Patent & Trademark Office is restricted to authorized employees and contractors only.

Form PTO-436A (Rev. 8/92)

(FAC

Affymetrix v. Illumina
C.A. No 04-901 JFF
Trial Exhibit
DTX 4

08/327525

Attorney Docket No. 16528X-82/1091

PATENT APPLICATION

COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION

Inventors:

MARK S. CHEE, who resides at
3199 Waverley Street
Palo Alto, California 94306
a citizen of Australia;

CHUNWEI WANG, who resides at
20350 Stevens Creek Boulevard #307
Cupertino, California 95014
a citizen of the People's Republic of China;

LUIS C. JEVONS, who resides at
701 Ramona Avenue
Sunnyvale, California 94087
a citizen of the United States of America;

DEREK H. BERNHART, who resides at
111 Seale Avenue
Palo Alto, California, 94301
a citizen of the United States of America; and

ROBERT J. LIPSHUTZ, who resides at
970 Palo Alto Avenue
Palo Alto, California, 94301
a citizen of the United States of America.

Assignee:

AFFYMAX TECHNOLOGIES N.V., located at
De Ruyderkade 62,
Curacao, Netherlands Antilles
a Netherlands Antilles corporation

Entity:

Large

TOWNSEND and TOWNSEND KHOURIE and CREW
Steuart Street Tower, 20th Floor
One Market
San Francisco, California 94105
(415) 326-2400

17

A/No FEE

08/327525

PATENT

Attorney Docket No. 16528X-82/1091

COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION

COPYRIGHT NOTICE

MICROFICHE APPENDIX

Microfiche Appendices A to E comprising five (5) sheets, totaling 272 frames are included herewith.

GOVERNMENT RIGHTS NOTICE

Portions of the material in this specification arose in the course of or under contract nos. 92ER81275 (SBIR) between Affymetrix, Inc. and the Department of Energy and/or H600813-1, -2 between Affymetrix, Inc. and the National Institutes of Health.

BACKGROUND OF THE INVENTION

The present invention relates to the field of computer systems. More specifically, the present invention relates to computer systems for visualizing biological sequences, as well as for evaluating and comparing biological sequences.

Devices and computer systems for forming and using arrays of materials on a substrate are known. For example, PCT application WO92/10588, incorporated herein by reference for all purposes, describes techniques for sequencing or sequence checking nucleic acids and other materials. Arrays for performing these operations may be formed in arrays according to the methods of, for example, the pioneering

techniques disclosed in U.S. Patent No. 5,143,854 and U.S. Patent Application No. 08/249,188, both incorporated herein by reference for all purposes.

According to one aspect of the techniques described therein, an array of nucleic acid probes is fabricated at known locations on a chip or substrate. A fluorescently labeled nucleic acid is then brought into contact with the chip and a scanner generates an image file indicating the locations where the labeled nucleic acids bound to the chip. Based upon the identities of the probes at these locations, it becomes possible to extract information such as the monomer sequence of DNA or RNA. Such systems have been used to form, for example, arrays of DNA that may be used to study and detect mutations relevant to cystic fibrosis, the P53 gene (relevant to certain cancers), HIV, and other genetic characteristics.

Improved computer systems and methods are needed to evaluate, analyze, and process the vast amount of information now used and made available by these pioneering technologies.

SUMMARY OF THE INVENTION

An improved computer-aided system for visualizing and determining the sequence of nucleic acids is disclosed. The computer system provides, among other things, improved methods of analyzing fluorescent image files of a chip containing hybridized nucleic acid probes in order to call bases in sample nucleic acid sequences.

According to one aspect of the invention, a computer system is used to identify an unknown base in a sample nucleic acid sequence by the steps of:

- inputting multiple probe intensities, each of the probe intensities being associated with a probe;
- the computer system comparing the multiple probe intensities where each of the probe intensities is substantially proportional to a probe hybridizing with at least one sequence; and

calling the unknown base according to the comparison of the multiple probe intensities.

According to one specific aspect of the invention, a higher probe intensity is compared to a lower probe intensity to call the unknown base. According to another specific aspect of the invention, probe intensities of a sample sequence are compared to probe intensities of a reference sequence. According to yet another specific aspect of the invention, probe intensities of a sample sequence are compared to statistics about probe intensities of a reference sequence from multiple experiments.

According to another aspect of the invention, a method is disclosed of processing reference and sample nucleic acid sequences to reduce the variations between the experiments by the steps of:

- providing a plurality of nucleic acid probes;
- labeling the reference nucleic acid sequence with a first marker;
- labeling the sample nucleic acid sequence with a second marker; and

hybridizing the labeled reference and sample nucleic acid sequences at the same time.

According to yet another aspect of the invention, a computer system is used for comparative analysis and visualization of multiple sequences by the steps of:

- displaying at least one reference sequence in a first area on a display device; and
- displaying at least one sample sequence in a second area on said display device;

whereby a user is capable of visually comparing the multiple sequences.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates an overall system for forming and analyzing arrays of biological materials such as DNA or RNA;

Fig. 2A is an illustration of the software for the overall system; Fig. 2B illustrates the global layout of a chip formed in the overall system; and Fig. 2C illustrates conceptually the binding of probes on chips;

Fig. 3 illustrates the high level flow of the intensity ratio method;

Fig. 4A illustrates the high level flow of one implementation of the reference method and Fig. 4B shows an analysis table for use with the reference method;

Fig. 5A illustrates the high level flow of another implementation of the reference method; Fig. 5B shows a data table for use with the reference method; Fig. 5C shows a graph of the normalized sample base intensities minus the normalized reference base intensities; and Fig. 5D shows other graphs of data in the data table;

Fig. 6 illustrates the high level flow of the statistical method;

Fig. 7 illustrates the pooling processing of a reference and sample nucleic acid sequence;

Fig. 8 illustrates the main screen and the associated pull down menus for comparative analysis and visualization of multiple experiments;

Fig. 9 illustrates an intensity graph window for a selected base;

Fig. 10 illustrates multiple intensity graph windows for selected bases;

Fig. 11 illustrates the intensity ratio method correctly calling a mutation in solutions with varying concentrations;

Fig. 12 illustrates the reference method correctly calling a mutant base where the intensity ratio method incorrectly called the mutant base; and

Fig. 13 illustrates the output of the ViewSeq™ program with four pretreatment samples and four posttreatment samples.

DESCRIPTION OF THE PREFERRED EMBODIMENT

CONTENTS

I. General

The present invention provides methods of analyzing hybridization intensity files for a chip containing hybridized nucleic acid probes. In a representative embodiment, the files represent fluorescence data from a biological array, but the files may also represent other data such as radioactive intensity data. For purposes of illustration, the present invention is described as being part of a computer system that designs a chip mask, synthesizes the probes on the chip, labels the nucleic acids, and scans the hybridized nucleic acid probes. Such a system is fully described in U.S. Patent Application No. 08/249,188 which has been incorporated by reference for all purposes. However, the present invention may be used separately from the overall system for analyzing data generated by such systems.

Fig. 1 illustrates a computerized system for forming and analyzing arrays of biological materials such as RNA or DNA. A computer 100 is used to design arrays of biological polymers such as RNA or DNA. The computer 100 may be, for example, an appropriately programmed Sun Workstation or personal computer or workstation, such as an IBM PC equivalent, including appropriate memory and a CPU. The computer system 100 obtains inputs from a user regarding characteristics of a gene of interest, and other inputs regarding the desired features of the array. Optionally, the computer system may obtain information regarding a specific

genetic sequence of interest from an external or internal database 102 such as GenBank. The output of the computer system 100 is a set of chip design computer files 104 in the form of, for example, a switch matrix, as described in PCT application WO 92/10092, and other associated computer files.

The chip design files are provided to a system 106 that designs the lithographic masks used in the fabrication of arrays of molecules such as DNA. The system or process 106 may include the hardware necessary to manufacture masks 110 and also the necessary computer hardware and software 108 necessary to lay the mask patterns out on the mask in an efficient manner. As with the other features in Fig. 1, such equipment may or may not be located at the same physical site, but is shown together for ease of illustration in Fig. 1. The system 106 generates masks 110 or other synthesis patterns such as chrome-on-glass masks for use in the fabrication of polymer arrays.

The masks 110, as well as selected information relating to the design of the chips from system 100, are used in a synthesis system 112. Synthesis system 112 includes the necessary hardware and software used to fabricate arrays of polymers on a substrate or chip 114. For example, synthesizer 112 includes a light source 116 and a chemical flow cell 118 on which the substrate or chip 114 is placed. Mask 110 is placed between the light source and the substrate/chip, and the two are translated relative to each other at appropriate times for deprotection of selected regions of the chip. Selected chemical reagents are directed through flow cell 118 for coupling to deprotected regions, as well as for washing and other operations. All operations are preferably directed by an appropriately programmed computer 119, which may or may not be the same computer as the computer(s) used in mask design and mask making.

The substrates fabricated by synthesis system 112 are optionally diced into smaller chips and exposed to marked receptors. The receptors may or may not be complementary to one or more of the molecules on the substrate. The receptors are marked with a label such as a fluorescein label (indicated

by an asterisk in Fig. 1) and placed in scanning system 120. Scanning system 120 again operates under the direction of an appropriately programmed digital computer 122, which also may or may not be the same computer as the computers used in synthesis, mask making, and mask design. The scanner 120 includes a detection device 124 such as a confocal microscope or CCD (charge-coupled device) that is used to detect the ~~location~~ locations where labeled receptor (*) has bound to the substrate. The output of scanner 120 is an image file(s) 124 indicating, in the case of fluorescein labeled receptor, the fluorescence intensity (photon counts or other related measurements, such as voltage) as a function of position on the substrate. Since higher photon counts will be observed where the labeled receptor has bound more strongly to the array of polymers, and since the monomer sequence of the polymers on the substrate is known as a function of position, it becomes possible to determine the sequence(s) of polymer(s) on the substrate that are complementary to the receptor.

The image file 124 is provided as input to an analysis system 126 that incorporates the visualization and analysis methods of the present invention. Again, the analysis system may be any one of a wide variety of computer system(s), but in a preferred embodiment the analysis system is based on a Sun Workstation or equivalent. The present invention provides various methods of analyzing the chip design files and the image files, providing appropriate output 128. The present invention may further be used to identify specific mutations in a receptor such as DNA or RNA.

Fig. 2A provides a simplified illustration of the overall software system used in the operation of one embodiment of the invention. As shown in Fig. 2A, the system first identifies the genetic sequence(s) or targets that would be of interest in a particular analysis at step 202. The sequences of interest may, for example, be normal or mutant portions of a gene, genes that identify heredity, or provide forensic information. Sequence selection may be provided via manual input of text files or may be from external sources such as GenBank. At step 204 the system evaluates the gene to

determine or assist the user in determining which probes would be desirable on the chip, and provides an appropriate "layout" on the chip for the probes. A wild-type probe is a probe that will ideally hybridize with the gene of interest and thus a wild-type gene (also called the chip wild-type) would ideally hybridize with all the wild-type probes on the chip. The layout implements desired characteristics such as arrangement on the chip that permits "reading" of genetic sequence and/or minimization of edge effects, ease of synthesis, and the like.

Fig. 2B illustrates the global layout of a chip. Chip 114 is composed of multiple units where each unit may contain different tilings for the chip wild-type sequence. Unit 1 is shown in greater detail and shows that each unit is composed of multiple cells which are areas on the chip that may contain probes. Conceptually, each unit is composed of multiple sets of related cells. As used herein, the term cell refers to a region on a substrate that contains many copies of a molecule or molecules of interest. Each unit is composed of multiple cells that may be placed in rows and columns. In one embodiment, a set of five related cells includes the following: a wild-type cell 220, "mutation" cells 222, and a "blank" cell 224. Cell 220 contains a wild-type probe that is the complement of a portion of the wild-type sequence. Cells 222 contain "mutation" probes for the wild-type sequence. For example, if the wild-type probe is 3'-ACGT, the probes 3'-ACAT, 3'-ACCT, 3'-ACGT, and 3'-ACTT may be the "mutation" probes. Cell 224 is the "blank" cell because it contains no probes (also called the "blank" probe). As the blank cell contains no probes, labeled receptors should not bind to the chip in this area. Thus, the blank cell provides an area that can be used to measure the background intensity.

In one embodiment, numerous tiling processes are available including sequence tiling, block tiling, and opt-tiling. ~~Of course~~ a wide range of layout strategies may be used according to the invention herein, without departing from the scope of the invention. For example, the probes may be tiled on a substrate in an apparently random fashion where a

computer system is utilized to keep track of the probe locations and correlate the data obtained from the substrate.

Opt-tiling is the process of tiling additional probes for suspected mutations. As a simple example of opt-tiling, suppose the wild-type target sequence is 5'-ACGTATGCA-3' and it is suspected that a mutant sequence has a possible T base mutation at the underlined base position. Suppose further that the chip will be synthesized with a "4x3" tiling strategy, meaning that probes of four monomers are used and that the monomers in position 3, counting left to right, of the probe are varied.

In opt-tiling, extra probes are tiled for each suspected mutation. The extra probes are tiled as if the mutation base is a wild-type base. The following shows the probes that may be generated for this example:

Table 1

Probe Sequences (From 3'-end)

4x3 Opt-Tiling

Wild	TGCA	GCAT	CATA	ATAC	TACG
A sub.	TGAA	GCAT	CAAA	ATAC	TAAG
C sub.	TGCA	GCCT	CACA	ATCC	TACG
G sub.	TGGA	GCGT	CAGA	ATGC	TAGG
T sub.	TGTA	GCTT	CATA	ATTC	TATG
Wild	TGCA	GCAA	CAAA	AAAC	AACG
A sub.	TGAA	GCAA	CAAA	AAAC	AAAG
C sub.	TGCA	GCCA	CACA	AACC	AACG
G sub.	TGGA	GCGA	CAGA	AAGC	AAGG
T sub.	TGTA	GCTA	CATA	AATC	AATG

In the first "chip" above, the top row of the probes (along with one probe below each of the four wild-type probes) should bind to the target DNA sequence. However, if the target sequence has a T base mutation as suspected, the labeled mutant sequence will not bind that strongly to the probes in the columns around column 3. For example, the mutant receptor that could bind with the probes in column 2 is 5'-CGTT which may not bind that strongly to any of the probes in column 2

because there are T bases at the ends of the receptor and probes (i.e., not complementary). This often results in a relatively dark scanned area around a mutation.

Opt-tiling ~~provides~~ generates the second "chip" above ~~which~~ to handle ~~treats~~ the suspected mutation as ~~the~~ a wild-type base. Thus, the mutant receptor 5'-CGTT should bind strongly to the wild-type probe of column 2 (along with one probe below) and the mutation can be further detected.

Again referring to Fig. 2A, at step 206 the masks for the synthesis are designed. At step 208 the software utilizes the mask design and layout information to make the DNA or other polymer chips. This software 208 will control ~~control, among other things,~~ relative translation of a substrate and the mask, the flow of desired reagents through a flow cell, the synthesis temperature of the flow cell, and other parameters. At step 210, another piece of software is used in scanning a chip thus synthesized and exposed to a labeled receptor. The software controls the scanning of the chip, and stores the data thus obtained in a file that may later be utilized to extract sequence information.

At step 212 a computer system according to the present invention utilizes the layout information and the fluorescence information to evaluate the hybridized nucleic acid probes on the chip. Among the important pieces of information obtained from DNA chips are the identification of mutant receptors and determination of genetic sequence of a particular receptor.

Fig. 2C illustrates the binding of a particular target DNA to an array of DNA probes 114. As shown in this simple example, the following probes are formed in the array (only one probe is shown for the wild-type probe):

```
3'-AGAACGT
   AGACCGT
   AGAGCGT
   AGATCGT
     .
     .
     .
```

As shown, the set of probes differ by only one base so the probes are designed to determine the identity of the base at that ~~location~~ position in the nucleic acid sequence.

When a fluorescein-labeled (or ~~other~~ otherwise marked) target with the sequence 5'-TCTTGCA is exposed to the array, it is complementary only to the probe 3'-AGAACGT, and fluorescein will be primarily found on the surface of the chip where 3'-AGAACGT is located. Thus, for each set of probes that differ by only one base, the image file will contain four fluorescence intensities, one for each probe. Each fluorescence intensity can therefore be associated with the base of each probe that is different from the other probes. Additionally, the image file will contain a "blank" cell which can be used as the fluorescence intensity of the background. By analyzing the five fluorescence intensities associated with a specific base location, it becomes possible to extract sequence information from such arrays using the methods of the invention disclosed herein.

The present invention calls bases by assigning the bases the following codes:

Code	Group	Meaning
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T(U)	Thymine (Uracil)
M	A or C	aMino
R	A or G	puRine
W	A or T(U)	Weak interaction (2 H bonds)
Y	C or T(U)	pYrimidine
S	C or G	Strong interaction (3 H bonds)
K	G or T(U)	Keto
V	A, C or G	not T(U)
H	A, C or T(U)	not G
D	A, G or T(U)	not C
B	C, G or T(U)	not A
N	A, C, G, or T(U)	Insufficient intensity to call
X	A, C, G, or T(U)	Insufficient discrimination to call

Most of the codes conform to the IUPAC standard. However, code N has been redefined and code X has been added.

II. Intensity Ratio Method

The intensity ratio method is a method of calling bases in a sample nucleic acid sequence. The intensity ratio method is most accurate when there is good discrimination between the fluorescence intensities of hybrid matches and hybrid mismatches. If there is insufficient discrimination, the intensity ratio method assigns a corresponding ambiguity code to the unknown base.

For simplicity, the intensity ratio method will be described as being used to identify one unknown base in a sample nucleic acid sequence. In practice, the method is used to identify many or all the bases in a nucleic acid sequence.

The unknown base will be identified by evaluation of up to four mutation probes and a "blank" cell, which is a location where a labeled receptor should not bind to the chip since no probe is present. For example, suppose a DNA sequence of interest or target sequence contains the sequence 5'-AGAACCTGC-3' with a possible mutation at the underlined base position. Suppose that 5-mer probes are to be synthesized for the target sequence. A representative wild-type probe of 5'-TTGGA is complementary to the region of the sequence around the possible mutation. The "mutation" probes will be the same as the wild-type probe except for a different base at the third position as follows: 3'-TTAGA, 3'-TTCGA, 3'-TTGGA, and 3'-TTTGA.

If the fluorescently marked sample sequence is exposed to the above four mutation probes, the intensity should be highest for the probe that binds most strongly to the sample sequence. Therefore, if the probe 3'-TTTGA shows the highest intensity, the unknown base in the sample will generally be called an A mutation because the probes are complementary to the sample sequence.

The mutation probes are identical to the wild-type probes except that they each contain one of the four A, C, G, or T "mutations" for the unknown base. Although one of the

"mutation" probes will ~~optimally~~ be identical to the wild-type probe, such redundant probes are intentionally synthesized for quality control and design consistency.

The identity of the unknown base is preferably determined by evaluating the relative fluorescence intensities of up to four of the mutation probes, and the "blank" cell. Because each mutation probe is identifiable by the mutation base, a mutation probe's intensity will be referred to as the "base intensity" of the mutation base.

As a simple example of the intensity ratio method, suppose a gene of interest (target) is an HIV protease gene with the sequence 5'-ATGTGGACAGTTGTA-3' (SEQ ID NO:1). Suppose further that a sample sequence is suspected to have the same sequence as the target sequence except for a mutation of base C to base T at the underlined base position. Although hundreds of probes may be synthesized on the chip, the complementary mutation probes synthesized to detect a mutation in the sample sequence at the suspected mutation position may be as follows:

```
3'-TATC
3'-TCTC
3'-TGTC   (wild-type)
3'-TTTC
```

The mutation probe 3'-TGTC is also the wild-type probe as it should bind most strongly with the target sequence.

After the sample sequence is labeled, hybridized on the chip, and scanned, suppose the following fluorescence intensities were obtained:

```
3'-TATC   -> 45
3'-TCTC   ->  8
3'-TGTC   -> 32
3'-TTTC   -> 12
```

where the intensity is measured by the photon count detected by the scanner. The "blank" cell had a fluorescence intensity of 2. The photon counts in the examples herein are representative (not actual data) and provided for illustration purposes. In practice, the actual photon counts will vary greatly depending on the experiment parameters and the scanner utilized.

Although each fluorescence intensity is from a probe, the probes may be characterized by their unique

mutation base so the bases may be said to have the following intensities:

A -> 45
C -> 8
G -> 32
T -> 12

Thus, base A will be described as having an intensity of 45, which corresponds to the intensity of the mutation probe with the mutation base A.

Initially, each mutation base intensity is reduced by the background or "blank" cell intensity. This is done as follows:

A -> 45 - 2 = 43
C -> 8 - 2 = 6
G -> 32 - 2 = 30
T -> 12 - 2 = 10

Then, the base intensities are sorted in descending order of intensity. The above bases would be sorted as follows:

A -> 43
G -> 30
T -> 10
C -> 6

Next, the highest intensity base is compared to the second highest intensity base. Thus, the ratio of the intensity of base A to the intensity of base G is calculated as follows: A:G = 43 / 30 = 1.4. The ratio A:G is then compared to a predetermined ratio cutoff, which is a number that specifies the ratio required to identify the unknown base. For example, if the ratio cutoff is 1.2, the ratio A:G is greater than the ratio cutoff (1.4 > 1.2) and the unknown base is called by the mutation probe containing the mutation A. As probes are complementary to the sample sequence, the sample sequence is called as having a mutation T, resulting in a called sample sequence of 5'-ATGTGGATAGTTGTA-3' (SEQ ID NO:2).

As another example, suppose everything else is the same as in the previous example except that the sorted background adjusted intensities were as follows:

C -> 42
A -> 40
G -> 10
T -> 8

The ratio of the highest intensity base to the second highest intensity base (C:A) is 1.05. Because this ratio is not greater than the ratio cutoff of 1.2, the unknown base will be called as being ambiguously one of two or more bases as follows.

The second highest intensity base is then compared to the third highest base. The ratio of A:G is 4. The ratio of A:G is then compared to the ratio cutoff of 1.2. As the ratio A:G is greater than the ratio cutoff (4 > 1.2), the unknown base is called by the mutation probes containing the mutations C or A. As probes are complementary to the sample sequence, the sample sequence is called as having either a mutation G or T, resulting in a sample sequence of 5'-ATGTGGAKAGTTGTA-3' (SEQ ID NO:3) where K is the IUPAC code for G or T(U).

The ratio cutoff in the previous examples was equal to 1.2. However, the ratio cutoff will generally need to be adjusted to produce optimal results for the specific chip design and wild-type target. Also, although the ratio cutoff used has been the same for each ratio comparison, the ratio cutoff may vary depending on whether the ratio comparisons involve the highest, second highest, third highest, etc. intensity base.

Fig. 3 illustrates the high level flow of the intensity ratio method. At step 302 the four base intensities are adjusted by subtracting the background or "blank" cell intensity from each base intensity. Preferably, if a base intensity is then less than or equal to zero, the base intensity is set equal to a small positive number to prevent division by zero or negative numbers in future calculations.

At step 304 the base intensities are sorted by intensity. Each base is then associated with a number from 1 to 4. The base with the highest intensity is 1, second highest 2, third highest 3, and fourth highest 4. Thus, the intensity of base 1 ≥ base 2 ≥ base 3 ≥ base 4.

At step 306 the highest intensity base (base 1) is checked to see if it has sufficient intensity to call the unknown base. The intensity is checked by determining if the intensity of base 1 is greater than a predetermined background

difference cutoff. The background difference cutoff is a number that specifies the intensity a base intensity must be over the background intensity in order to correctly call the unknown base. Thus, the background adjusted base intensity must be greater than the background difference cutoff or the unknown is not callable.

If the intensity of base 1 is not greater than the background difference cutoff, the unknown base is assigned the code N (insufficient intensity) as shown at step 308. Otherwise, the ratio of the intensity of base 1 to base 2 is calculated as shown at step 310.

At step 312 the ratio of ~~intensity~~ intensities of ~~base~~ bases 1:2 is compared to the ratio cutoff. If the ratio 1:2 is greater than the ratio cutoff, the unknown base is called as the complement of the highest intensity base (base 1) as shown at step 314. Otherwise, the ratio of the intensity of base 2 to base 3 is calculated as shown at step 316.

At step 318 the ratio of ~~intensity~~ intensities of ~~base~~ bases 2:3 is compared to the ratio cutoff. If the ratio 2:3 is greater than the ratio cutoff, the unknown base is called as being an ambiguity code specifying the complements of the highest or second highest intensity bases (base 1 or 2) as shown at step 320. Otherwise, the ratio of the intensity of base 3 to base 4 is calculated as shown at step 322.

At step 324 the ratio of ~~intensity~~ intensities of ~~base~~ bases 3:4 is compared to the ratio cutoff. If the ratio 3:4 is greater than the ratio cutoff, the unknown base is called as being an ambiguity code specifying the complements of the highest, second highest, or third highest bases (base 1, 2 or 3) as shown at step 326. Otherwise, the unknown base is assigned the code X (insufficient discrimination) as shown at step 328.

The advantage of the intensity ratio method is that it is very accurate when there is good discrimination between the fluorescence intensities of hybrid matches and hybrid mismatches. However, if the base corresponding to a correct hybrid gives a lower intensity than a mismatch (e.g., as a result of cross-hybridization), incorrect identification of the base will result. For this reason, however, the method is

useful for comparative assessment of hybridization quality and as an indicator of sequence-specific problem spots. For example, the intensity ratio method has been used to determine that ambiguities and miscalls tend to be very different from sequence to sequence, and reflect predominantly the composition and repetitiveness of the sequence. It has also been used to assess improvements obtained by varying hybridization conditions, sample preparation, and post-hybridization treatments (e.g., RNase treatment).

III. Reference Method

The reference method is a method of calling bases in a sample nucleic acid sequence. The reference method depends very little on discrimination between the fluorescence intensities of hybrid matches and hybrid mismatches, and therefore is much less sensitive to cross-hybridization. The method compares the probe intensities of a reference sequence to the probe intensities of a sample sequence. Any significant changes are flagged as possible mutations. There are two implementations of the reference method disclosed herein.

For simplicity, the reference method will be described as being used to identify one unknown base in a sample nucleic acid sequence. In practice, the method is used to identify many or all the bases in a nucleic acid sequence.

The unknown base will be called by comparing the probe intensities of a reference sequence to the probe intensities of a sample sequence. Preferably, the probe intensities of the reference sequence and the sample sequence are from chips having the same chip wild-type. However, the reference sequence may or may not be exactly the same as the chip wild-type, as it may have mutations.

The bases at the same position in the reference and sample sequences will each be associated with up to four mutation probes and a "blank" cell. The unknown base in the sample sequence is called by comparing probe intensities of the sample sequence to probe intensities of the reference sequence. For example, suppose the chip wild-type contains

the sequence 5'-AGACCTTGC-3' and it is suspected that the sample has a possible mutation at the underlined base position, which is the unknown base that will be called by the reference method. The "mutation" probes for the sample sequence may be as follows: 3'-GAAA, 3'-GCAA, 3'-GGAA, and 3'-GTAA, where 3'-GGAA is the wild-type probe.

Suppose further that a reference sequence, which differs from the chip wild-type by one base mutation, has the sequence 5'-AGACATTGC-3' where the mutation base is underlined. The "mutation" probes for the reference sequence may be as follows: 3'-TGAAA, 3'-TGCAA, 3'-TGGAA, and 3'-TGTAA, where 3'-TGTAA is the reference wild-type probe since the reference sequence is known. Although generally the sample and reference sequences were tiled with the same chip wild-type, this is not required, and the tiling methods do not have to be identical as shown in the example. Thus, the unknown base will be called by comparing the "mutation" probes of the sample sequence to the "mutation" probes of the reference sequence. As before, because each mutation probe is identifiable by the mutation base, the mutation probes' intensities will be referred to the "base intensities" of their respective mutation bases.

As a simple example of one implementation of the reference method, suppose a gene of interest (target) has the sequence 5'-AAAACTGAAAA-3' (SEQ ID NO:4). Suppose a reference sequence has the sequence 5'-AAAACCGAAAA-3' (SEQ ID NO:5), which differs from the target sequence by the underlined base. The reference sequence is marked and exposed to probes on a chip with the target sequence being the chip wild-type. Suppose further that a sample sequence is suspected to have the same sequence as the target sequence except for a mutation at the underlined base position in 5'-AAAACTGAAAA-3' (SEQ ID NO:6). The sample sequence is also marked and exposed to probes on a chip with the target sequence being the chip wild-type. After hybridization and scanning, the following probe intensities (not actual data) were found for the respective complementary probes:

Reference		Sample	
3'-TGAC	-> 12	3'-GACT	-> 11
3'-TGCC	-> 9	3'-GCCT	-> 30
3'-TGGC	-> 80	3'-GGCT	-> 60
3'-TGTC	-> 15	3'-GTCT	-> 6

Although each fluorescence intensity is from a probe, the probes may be identified by their unique mutation base so the bases may be said to have the following intensities:

Reference		Sample	
A	-> 12	A	-> 11
C	-> 9	C	-> 30
G	-> 80	G	-> 60
T	-> 15	T	-> 6

Thus, base A of the reference sequence will be described as having an intensity of 12, which corresponds to the intensity of the mutation probe with the mutation base A. The reference method will now be described as calling the unknown base in the sample sequence by using these intensities.

Fig. 4A illustrates the high level flow of one implementation of the reference method. For illustration purposes, the reference method is described as filling in the columns (identified by the numbers along the bottom) of the analysis table shown in Fig. 4B. However, the generation of an analysis table is not necessary to practice the method. The analysis table is shown to aid the reader in understanding the method.

At step 402 the four base intensities of the reference and sample sequences are adjusted by subtracting the background or "blank" cell intensity from each base intensity. Each set of "mutation" probes has an associated "blank" cell. Suppose that the reference "blank" cell intensity is 1 and the sample "blank" cell intensity is 2. The base intensities are then background subtracted as follows:

Reference		Sample	
A	-> 12 - 1 = 11	A	-> 11 - 2 = 9
C	-> 9 - 1 = 8	C	-> 30 - 2 = 28
G	-> 80 - 1 = 79	G	-> 60 - 2 = 58
T	-> 15 - 1 = 14	T	-> 6 - 2 = 4

Preferably, if a base intensity is then less than or equal to zero, the base intensity is set equal to a small positive number to prevent division by zero or negative numbers in future calculations.

a For identification, the position of ~~the bases~~ each base of interest in the reference and sample sequences is placed in column 1 of the analysis table. Also, since the reference sequence is a known sequence, the base at this position is known and is referred to as the reference wild-type. The reference wild-type is placed in column 2 of the analysis table, which is C for this example.

At step 404 the base intensity associated with the reference wild-type (column 2 of the analysis table) is checked to see if it has sufficient intensity to call the unknown base. In this example, the reference wild-type is C. However, the base intensity associated with the wild-type is the G base intensity, which is 79 in this example. This is because the base intensities actually represent the complementary "mutation" probes. The G base intensity is checked by determining if its intensity is greater than a predetermined background difference cutoff. The background difference cutoff is a number that specifies the intensity the base intensities must be above the background intensity in order to correctly call the unknown base. Thus, the base intensity associated with the reference wild-type must be greater than the background difference cutoff or the unknown base is not callable.

If the background difference cutoff is 5, the base intensity associated with the reference wild-type has sufficient intensity (79 > 5) so a P (pass) is placed in column 3 of the analysis table as shown at step 406. Otherwise, at step 407 an F (fail) is placed in column 3 of the analysis table.

At step 408 the ratio of the base intensity associated with the reference wild-type to each of the possible bases are calculated. The ratio of the base intensity associated with the reference wild-type to itself will be 1 and the other ratios will usually be greater than 1. The base intensity associated with the reference wild-type is G so the following ratios are calculated:

G:A -> 79 / 11 = 7.2
G:C -> 79 / 8 = 9.9
G:G -> 79 / 79 = 1.0
G:T -> 79 / 14 = 5.6

These ratios are placed in columns 4 through 7 of the analysis table, respectively.

At step 410 the highest base intensity associated with the sample sequence is checked to see if it has sufficient intensity to call the unknown base. The highest base intensity is checked by determining if the intensity is greater than the background difference cutoff. Thus, the highest base intensity must be greater than the background difference cutoff or the unknown base is not callable.

Again, if the background difference cutoff is 5, the highest base intensity, which is G in this example, has sufficient intensity (58 > 5) so a P (pass) is placed in column 8 of the analysis table as shown at step 412. Otherwise, at step 413 an F (fail) is placed in column 8 of the analysis table.

At step 414 the ratios of the highest base intensity of the sample to each of the possible bases are calculated. The ratio of the highest base intensity to itself will be 1 and the other ratios will usually be greater than 1. Thus, the highest base intensity is G so the following ratios are calculated:

G:A -> 58 / 9 = 6.4
G:C -> 58 / 28 = 2.3
G:G -> 58 / 58 = 1.0
G:T -> 58 / 4 = 14.5

These ratios are placed in columns 9 through 12 of the analysis table, respectively.

At step 416 if both the reference and sample sequence probes failed to have sufficient intensity to call the unknown base, meaning there is an 'F' in columns 3 and 8 of the analysis table, the unknown base is assigned the code N (insufficient intensity) as shown at step 418. An 'N' is placed in column 17 of the analysis table. Additionally, a confidence code of 9 is placed in column 18 of the analysis table where the confidence codes have the following meanings:

Code	Meaning
0	Probable reference wild-type
1	Probable mutation
2	Reference sufficient intensity, insufficient intensity in sample suggests possible mutation
3	Borderline differences, unknown base ambiguous
4	Sample sufficient intensity, insufficient intensity in reference to allow comparison
5-8	Currently unassigned
9	Insufficient intensity in reference and sample, no interpretation possible

The confidence codes are useful for indicating to the user the resulting analysis of the reference method.

At step 420 if only the reference sequence probes failed to have sufficient intensity to call the unknown base, meaning there is an 'F' in column 3 and a 'P' in column 8 of the analysis table, the unknown base is assigned the code N (insufficient intensity) as shown at step 422. An 'N' is placed in column 17 and a confidence code of 4 is placed in column 18 of the analysis table.

At step 424 if only the sample sequence probes failed to have sufficient intensity to call the unknown base, meaning there is a 'P' in column 3 and a 'F' in column 8 of the analysis table, the unknown base is assigned the code N (insufficient intensity) as shown at step 426. An 'N' is placed in column 17 and a confidence code of 2 is placed in column 18 of the analysis table.

In this example, both the reference and sample sequence probes have sufficient intensity to call the unknown base. At step 428 the ratios of the reference ratios to the sample ratios for each base type are calculated. Thus, the ratio A:A (column 4 to column 9) is placed in column 13 of the analysis table. The ratio C:C (column 5 to column 10) is placed in column 14 of the analysis table. The ratio G:G (column 6 to column 11) is placed in column 15 of the analysis table. Lastly, the ratio T:T (column 7 to column 12) is placed in column 16 of the analysis table. These ratios are calculated as follows:

A:A -> 7.2 / 6.4 = 1.1
C:C -> 9.9 / 2.3 = 4.3
G:G -> 1.0 / 1.0 = 1.0
T:T -> 5.6 / 14.5 = 0.4

The unknown base is called by comparing these ratios of ratios to two predetermined values as follows.

At step 430 if all the ratios of ratios (columns 13 to 16 of the analysis table) are less than a predetermined lower ratio cutoff, the unknown base is assigned the code of the reference wild-type as shown at step 432. Thus, the code for the reference wild-type (as shown in column 2) would be placed in column 17 and a confidence code of 0 would be placed in column 18 of the analysis table.

At step 434 if all the ratios of ratios are less than a predetermined upper ratio cutoff, the unknown base is assigned an ambiguity code that indicates the unknown base may be any one of the bases that has a complementary ratio of ratios greater than the lower ratio cutoff and less than the upper ratio cutoff as shown at step 436. Thus, if the ratio of ratios for A:A, C:C and G:G are all greater than the lower ratio cutoff and less than the upper ratio cutoff, the unknown base would be assigned the code B (meaning "not A"). This is because the ratios of ratios are complementary to their respective base as follows:

A:A -> T
C:C -> G
G:G -> C

so the unknown base would be called as being either C, G, or T, which is identified by the IUPAC code B. This ambiguity code would be placed in column 17 and a confidence code of 3 would be placed in column 18 of the analysis table.

At step 438 at least one of the ratios of ratios is greater than the upper ratio cutoff and the unknown base is called as the base complementary to the highest ratio of ratios. The code for the base complementary to the highest ratio of ratios would be placed in column 17 and a confidence code of 1 would be placed in column 18 of the analysis table.

Assume for the purposes of this example that the lower ratio cutoff is 1.5 and the upper ratio cutoff is 3. Again, the ratios of ratios are as follows:

A:A -> 1.1
C:C -> 4.3
G:G -> 1.0
T:T -> 0.4

As all the ratios of ratios are not less than the upper ratio cutoff, the unknown base is called the base complementary to the highest ratio of ratios. The highest ratio of ratios is C:C, which has a complementary base G. Thus, the unknown base is called G which is placed in column 17 and a confidence code of 1 is placed in column 18 of the analysis table.

The example shows how the unknown base in the sample nucleic acid sequence was correctly called as base G. Although the complementary "mutation" probe associated with the base G (3'-GCCT) did not have the highest fluorescence intensity, the unknown base was called as base G because the associated "mutation" probe had the highest ratio increase over the other "mutation" probes.

Fig. 5A illustrates the high level flow of another implementation of the reference method. As in the previous implementation, this implementation also compares the probe intensities of a reference sequence to the probe intensities of a sample sequence. However, this implementation differs conceptually from the previous implementation in that neighboring probe intensities are also analyzed, resulting in more accurate base calling.

As a simple example of this implementation of the reference method, suppose a reference sequence has a sequence of 5'-AAACCCAATCCACATCA-3' (SEQ ID NO:6) and a sample sequence has a sequence of 5'-AAACCCAGTCCACATCA-3' (SEQ ID NO:7), where the mutant base is underlined. Thus, there is a mutation of A to G. Suppose further that the reference and sample sequences are tiled on chips with the reference sequence being the chip wild-type. This implementation of the reference method will be described as identifying this mutation base.

For illustration purposes, this implementation of the reference method is described as filling in a data table (SEQ ID NO:6, SEQ ID NO: 28, and SEQ ID NO:29) shown in Fig. 5B. Although the data table contains more data than is required for this implementation, the portions of the data table that are produced by steps in Fig. 5A are shown

with the same reference numerals. The generation of a data table is not necessary, however, and is shown to aid the reader in understanding the method. The mutant base position is at position 241 in the reference and sample sequences, which is shown in bold in the data table.

At step 502 the base intensities of the reference and sample sequences are adjusted by subtracting the background or "blank" cell intensity from each base intensity. Preferably, if a base intensity is then less than or equal to zero, the base intensity is set equal to a small positive number to prevent division by zero or negative numbers. In the data table, data 502A is the background subtracted base intensities for the reference sequence and data 502B is the background subtracted base intensities for the sample sequence (also called the "mutant" sequence in the data table).

At step 504 the base intensity associated with the reference wild-type is checked to see if it has sufficient intensity to call the unknown base. In this example, the reference wild-type is base A at position 241. The base intensity associated with the reference wild-type is identified by a lower case "a" in the left hand column. Thus, the base intensities in the data table are not identified by their complements and the reference wild-type at the mutation position has an intensity of 385. The reference wild-type intensity of 385 is checked by determining if its intensity is greater than a predetermined background difference cutoff. The background difference cutoff is a number that specifies the intensity the base intensities must be over the background intensity in order to correctly call the unknown base. Thus, the base intensity associated with the reference wild-type must be greater than the background difference cutoff or the unknown base is not callable.

If the base intensity associated with the reference wild-type is not greater than the background difference cutoff, the wild-type sequence would fail to have sufficient intensity as shown at step 506. Otherwise, at step 508 the wild-type sequence would pass by having sufficient intensity.

At step 510 calculations are performed on the background subtracted base intensities of the reference sequence in order to "normalize" the intensities. Each position in the reference sequence has four background subtracted base intensities associated with it. The ratio of the intensity of each base to the sum of the intensities of the possible bases (all four) is calculated, resulting in four ratios, one for each base as shown in the data table. Thus, the following ratios would be calculated at each position in the reference sequence:

A ratio = A / (A + C + G + T)
C ratio = C / (A + C + G + T)
G ratio = G / (A + C + G + T)
T ratio = T / (A + C + G + T)

At position 241, A ratio would be the wild-type ratio. These ratios are generally calculated in order to "normalize" the intensity data as the photon counts may vary widely from experiment to experiment. Thus, the ratios provide a way of reconciling the intensity variations across experiments. Preferably, if the photon counts do not vary widely from experiment to experiment, the probe intensities do not need to be "normalized."

At step 512 the highest base intensity associated with the sample sequence is checked to see if it has sufficient intensity to call the unknown base. The intensity is checked by determining if the highest intensity sample base is greater than the background difference cutoff. If the intensity is not greater than the background difference cutoff, the sample sequence fails to have sufficient intensity as shown at step 514. Otherwise, at step 516 the sample sequence passes by having sufficient intensity.

At step 518 calculations are performed on the background subtracted base intensities of the sample sequence in order to "normalize" the intensities. Each position in the sample sequence has four background subtracted base intensities associated with it. The ratios of the intensity of each base to the sum of the intensities of the possible bases (all four) are calculated, resulting in four ratios, one for each base as shown in the data table.

At step 520 if either the reference or sample sequences failed to have sufficient intensity, the unknown base is assigned the code N (insufficient intensity) as shown at step 522.

At step 524 the normalized base intensity ratios of the reference sequence are subtracted from the normalized base intensity ratios of the sample sequence. Thus, at each position the following calculations are performed:

A Difference = Sample A Ratio - Reference A Ratio
C Difference = Sample C Ratio - Reference C Ratio
G Difference = Sample G Ratio - Reference G Ratio
T Difference = Sample T Ratio - Reference T Ratio

where the reference and sample ratios are calculated at steps 510 and 518, respectively. The base differences resulting from these calculations are shown in the data table.

At step 526 each position is checked to see if there is a base difference greater than an upper difference cutoff and a base difference lower than a lower difference cutoff. For example, Fig. 5C shows a graph the normalized sample base intensities minus the normalized reference base intensities. Suppose that the upper difference cutoff is 0.15 and the lower difference cutoff is -0.15 as shown by the horizontal lines in Fig. 5C. At the mutation position (labeled with a reference 0), the G difference is 0.28 which is greater than 0.15, the upper difference cutoff. Similarly, the A difference is -0.32 which is less than -0.15, the lower difference cutoff. As there is a base difference above the upper difference cutoff and a base difference below the lower difference cutoff, there may be mutation at this position.

If there is neither a base difference above the upper difference cutoff nor a base difference below the lower difference cutoff, the base at that position is assigned the code of the reference wild-type base as shown at step 528.

At step 530 the ratio of the highest background subtracted base intensity in the sample to the background subtracted reference wild-type base intensity is calculated. For example, at the mutation position 241 in the data table, the highest background subtracted base intensity in the sample is 571 (base G). The background subtracted reference wild-

type base intensity is 385 (base A). Thus, the ratio of 571:385 is calculated and results in 1.48 as shown in the data table.

At step 532 these ratios are compared to a ratio at a neighboring position. The ratio for the n^{th} position is subtracted from the ratio for the r^{th} position, where $r = n + 1$. For example, at the mutation position 241 in the data table, the ratio at position 242 (which equals 1.02) is subtracted from the ratio at position 241 (which equals 1.48). It has been found that a mutant can be confidently detected by analyzing the difference of these neighboring ratios.

Fig. 5D shows other graphs of data in the data table. Of particular importance is the graph identified as 532 because this is a graph of the calculations at step 532. The pattern shown in a box in graph 532 has been found to be characteristic of a mutation. Thus, if this pattern is detected, the base is called as the base (or bases) with a normalized difference greater than the upper difference cutoff as shown at step 536. For example, the pattern was detected and at step 526 it was shown that base G had a normalized difference of 0.28, which is greater than the upper difference cutoff of 0.15. Therefore, the base at position 241 in the sample sequence is called a base G, which is a mutation from the reference sequence (A to G).

If the pattern is not detected at step 534, the base at that position is assigned the code of the reference wild-type base as shown at step 538.

This second implementation of the reference method is preferable in some instances as it takes into account probe intensities of neighboring probes. The first implementation may not have detected the A to G mutation in this example.

The advantage of the reference method is that the correct base can be called even in the presence of significant levels of cross-hybridization, as long as ratios of intensities are fairly consistent from experiment to experiment. In practice, the number of miscalls and ambiguities is significantly reduced, while the number of

correct calls is actually increased, making the reference method very useful for identifying candidate mutations. The reference method has also been used to compare the reproducibility of experiments in terms of base calling.

IV. Statistical Method

The statistical method is a method of calling bases in a sample nucleic acid sequence. The statistical method utilizes the statistical variation across experiments to call the bases. Therefore, the statistical method is preferable when ~~good at calling bases if~~ data from multiple experiments is available and the data is fairly consistent ~~among~~ across the experiments. The method compares the probe intensities of a sample sequence to statistics of probe intensities of a reference sequence in multiple experiments.

For simplicity, the statistical method will be described as being used to identify one unknown base in a sample nucleic acid sequence. In practice, the method is used to identify many or all the bases in a nucleic acid sequence.

The unknown base will be called by comparing the probe intensities of a sample sequence to statistics on probe intensities of a reference sequence in multiple experiments. Generally, the probe intensities of the sample sequence and the reference sequence experiments are from chips having the same chip wild-type. However, the reference sequence may or may not be equal to the chip wild-type, as it may have mutations.

A base at the same position in the reference and sample sequences will be associated with up to four mutation probes and a "blank" cell. As before, because each mutation probe is identifiable by the mutation base, the mutation probes' intensities will be referred to as the "base intensities" of their respective mutation bases.

As a simple example of the statistical method, suppose a gene of interest (target) has the sequence 5'-AAAACTGAAAA-3' (SEQ ID NO:4). Suppose a reference sequence has the sequence 5'-AAAACCGAAAA-3' (SEQ ID NO:5), which differs from the target sequence by the underlined base. Suppose further that a sample sequence

is suspected to have the same sequence as the target sequence except for a T base mutation at the underlined base position in 5'-AAAACTGAAAA-3' (SEQ ID NO:4). Suppose that in multiple experiments the reference sequence is marked and exposed to probes on a chip. Suppose further the sample sequence is also marked and exposed to probes on a chip.

The following are complementary "mutation" probes that could be used for a reference experiment and the sample sequence:

Reference	Sample
3'-TGAC	3'-GACT
3'-TGCC	3'-GCCT
3'-TGGC	3'-GGCT
3'-TGTC	3'-GTCT

The "mutation" probes shown for the reference sequence may be from only one experiment, the other experiments may have different "mutation" probes, chip wild-types, tiling methods, and the like. Although each fluorescence intensity is from a probe, since the probes may be identified by their unique mutation bases, the probe intensities may be identified by their respective bases as follows:

Reference		Sample	
3'-TGAC	-> A	3'-GACT	-> A
3'-TGCC	-> C	3'-GCCT	-> C
3'-TGGC	-> G	3'-GGCT	-> G
3'-TGTC	-> T	3'-GTCT	-> T

Thus, base A of the reference sequence will be described as having an intensity which corresponds to the intensity of the mutation probe with the mutation base A. The statistical method will now be described as calling the unknown base in the sample sequence by using this example.

Fig. 6 illustrates the high level flow of the statistical method. At step 602 the four base intensities associated with the sample sequence and each of the multiple reference experiments are adjusted by subtracting the background or "blank" cell intensity from each base intensity. Preferably, if a base intensity is then less than or equal to zero, the base intensity is set equal to a small positive number to prevent division by zero or negative numbers.

At step 604 the intensities of the reference wild-type bases in the multiple experiments are checked to see if

they all have sufficient intensity to call the unknown base. The intensities are checked by determining if the intensity of the reference wild-type base of an experiment is greater than a predetermined background difference cutoff. The wild-type probe shown earlier for the reference sequence is 3'-TGGC, and thus the G base intensity is the wild-type base intensity. These steps are analogous to steps in the other two methods described herein.

If the intensity of any one of the reference wild-type bases is not greater than the background difference cutoff, the wild-type experiments fail to have sufficient intensity as shown at step 606. Otherwise, at step 608 the wild-type experiments pass by having sufficient intensity.

At step 610 calculations are performed on the background subtracted base intensities of each of the reference experiments in order to "normalize" the intensities. Each reference experiment has four background subtracted base intensities associated with it: one wild-type and three for the other possible bases. In this example, the G base intensity is the wild-type, the A, C, and T base intensities being the "other" intensities. The ratios of the intensity of each base to the sum of the intensities of the possible bases (all four) are calculated, giving one wild-type ratio and three "other" ratios. Thus, the following ratios would be calculated:

A ratio = A / (A + C + G + T)
C ratio = C / (A + C + G + T)
G ratio = G / (A + C + G + T)
T ratio = T / (A + C + G + T)

where G ratio is the wild-type ratio and A, C, and T ratios are the "other" ratios. These four ratios are calculated for each reference experiment. Thus if the number of reference experiments is n, there would be 4n ratios calculated. These ratios are generally calculated in order to "normalize" the intensity data, as the photon counts may vary widely from experiment to experiment. However, if the probe intensities do not vary widely from experiment to experiment, the probe intensities do not need to be "normalized."

At step 612 statistics are prepared for the ratios calculated for each of the reference experiments. As stated before, each reference experiment will be associated with one wild-type ratio and three "other" ratios. The mean and standard deviation are calculated for all the wild-type ratios. The mean and standard deviation are also calculated for each of the other ratios, resulting in three other means and standard deviations for each of the bases that is not the wild-type base. Therefore, the following would be calculated:

Mean and standard deviation of A ratios
Mean and standard deviation of C ratios
Mean and standard deviation of G ratios
Mean and standard deviation of T ratios

where the mean and standard deviation of the G ratios are also known as the wild-type mean and the wild-type standard deviation, respectively. The mean and standard deviation of the A, C, and T means and standard deviations are also known collectively as the "other" means and standard deviations.

Suppose that the preceding calculations produced the following data:

A ratios	->	mean = 0.16	std. dev. = 0.003
C ratios	->	mean = 0.03	std. dev. = 0.002
G ratios	->	mean = 0.71	std. dev. = 0.050
T ratios	->	mean = 0.11	std. dev. = 0.004

In one embodiment, the steps up to and including step 612 are performed in a preprocessing stage for the multiple wild-type experiments. The results of the preprocessing stage are stored in a file so that the reference calculations do not have to be repeatedly calculated, improving ~~which results in increased~~ performance. Microfiche Appendices C and D contain the programming code to perform the preprocessing stage.

At step 614 the highest base intensity associated with the sample sequence is checked to see if it has sufficient intensity to call the unknown base. The intensity is checked by determining if the highest intensity unknown base is greater than the background difference cutoff. If the intensity is not greater than the background difference cutoff, the sample sequence fails to have sufficient intensity

as shown at step 616. Otherwise, at step 618 the sample sequence passes by having sufficient intensity.

At step 620 calculations are performed on the four background subtracted intensities of the sample sequence. The ratio of the background subtracted intensity of each base to the sum of the background subtracted intensities of the possible bases (all four) are calculated, giving four ratios, one for each base. For consistency, the ratio associated with the reference wild-type base is called the wild-type ratio, with there being three "other" ratios. Thus, the following ratios are calculated:

```
A ratio = A / (A + C + G + T)
C ratio = C / (A + C + G + T)
G ratio = G / (A + C + G + T)
T ratio = T / (A + C + G + T)
```

where ratio G is the wild-type ratio and ratios A, C, and T are the "other" ratios.

Suppose the background subtracted intensities associated with the sample are as follows:

```
A    -> 310
C    ->  50
G    ->  26
T    -> 100
```

Then, the corresponding ratios would be as follows:

```
A ratio = 310 / (310 + 50 + 26 + 100) = 0.64
C ratio =  50 / (310 + 50 + 26 + 100) = 0.10
G ratio =  26 / (310 + 50 + 26 + 100) = 0.05
T ratio = 100 / (310 + 50 + 26 + 100) = 0.21
```

At step 622 if either the reference experiments or the sample sequence failed to have sufficient intensity, the unknown base is assigned the code N (insufficient intensity) as shown at step 624.

At step 626 the wild-type and "other" ratios associated with the sample sequence are compared to statistical expressions. The statistical expressions include four predetermined standard deviation cutoffs, one associated with each base. Thus, there is a standard deviation cutoff for each of the bases A, C, G, and T. The standard deviation cutoffs allow the unknown base to be called with higher precision because each standard deviation cutoff can be set to

a different value. Suppose the standard deviation cutoffs are set as follows:

A standard deviation cutoff -> 4
C standard deviation cutoff -> 2
G standard deviation cutoff -> 8
T standard deviation cutoff -> 4

The wild-type base ratio associated with the sample is compared to a corresponding statistical expression:

WT ratio ≥ WT mean - (WT std. dev. * WT base std. dev. cutoff)

where the WT base std. dev. cutoff is the standard deviation cutoff for the wild-type base. As the wild-type base is G, the above comparison solves to the following:

0.05 ≥ 0.71 - (0.050 * 8)
0.05 ≥ 0.31

which is not a true expression (0.05 is not greater than 0.31).

Each of the "other" ratios associated with the sample is compared to a corresponding statistical expression:

Other ratio > Other mean + (Other std. dev. * Other base std. dev. cutoff)

where the Other base std. dev. cutoff is the standard deviation cutoff for the particular "other" base. Thus, the above comparison solves to the following three expressions:

A -> 0.64 > 0.16 + (0.003 * 4)
0.64 > 0.17
C -> 0.10 > 0.03 + (0.002 * 2)
0.10 > 0.03
T -> 0.21 > 0.11 + (0.004 * 4)
0.21 > 0.13

which are all true expressions.

At step 628 if only the wild-type ratio of the sample sequence was greater than the statistical expression, the unknown base is assigned the code of the reference wild-type base as shown at step 630.

At step 632 if one or more of the "other" ratios of the sample sequence were greater than their respective statistical expressions, the unknown base is assigned an ambiguity code that indicates the unknown base may be any one of the complements of these bases, including the reference wild-type. In this example, the "other" ratios for A, C, and T were all greater than their corresponding statistical

expression. Thus, the unknown base would be called the complements of these bases, represented by the subset T, G, and A. Thus, the unknown base would be assigned the code D (meaning "not C").

If none of the ratios are greater than their respective statistical expressions, the unknown base is assigned the code X (insufficient discrimination) as shown at step 636.

The statistical method provides accurate base calling because it utilizes statistical data from multiple reference experiments to call the unknown base. The statistical method has also been used to implement confidence estimates and calling of mixed sequences.

V. Pooling Processing

The present invention provides pooling processing which is a method of processing reference and sample nucleic acid sequences together to reduce variations across individual experiments. In the representative embodiment discussed herein, the reference and sample nucleic acid sequences are labeled with different fluorescent markers emitting light at different wavelengths. However, the nucleic acids may be labeled with other types of markers including distinguishable radioactive markers.

After the reference and sample nucleic acid sequences are labeled with different color fluorescent markers, the labeled reference and sample nucleic acid sequences are then combined and processed together. An apparatus for detecting targets labeled with different markers is provided in U.S. Application No. 08/195,889 and is hereby incorporated by reference for all purposes.

Fig. 7 illustrates the pooling processing of a reference and sample nucleic acid sequence. At step 702 a reference nucleic acid sequence is marked with a fluorescent dye, such as a fluorescein. At step 704 a sample nucleic acid sequence is marked with a dye that, upon excitation, emits light ~~that~~ of a different wavelength than the fluorescent dye

of the reference sequence. For example, the sample nucleic acid sequence may be marked with rhodamine.

At step 706 the labeled reference sequence and the labeled sample sequence are combined. After this step, processing continues in the same manner as for only one labeled sequence. At step 708 the sequences are fragmented. The fragmented nucleic acid sequences are then hybridized on a chip containing probes as shown at step 710.

At step 712 a scanner generates image files that indicate the locations where the labeled nucleic acids bound to the chip. In general, the scanner generates an image file by focusing excitation light on the hybridized chip and detecting the fluorescent light that is emitted. The marker emitting the fluorescent light can be identified by the wavelength of the light. For example, the fluorescence peak of fluorescein is about 530 nm while that of a typical rhodamine dye is about 580 nm.

The scanner creates an image file for the data associated with each fluorescent marker, indicating the locations where the correspondingly labeled nucleic acid bound to the chip. Based upon an analysis of the fluorescence intensities and locations, it becomes possible to extract information such as the monomer sequence of DNA or RNA.

Pooling processing reduces variations across individual experiments because much of the test environment is common. Although pooling processing has been described as being used to improve the combined processing of reference and sample nucleic acid sequences, the process may also be used for two reference sequences, two sample sequences, or multiple sequences by utilizing multiple distinguishable markers.

VI. Comparative Analysis (ViewSeq™)

The present invention provides a method of comparative analysis and visualization of multiple experiments. The method allows the intensity ratio, reference, and statistical methods to be run on multiple datafiles simultaneously. This permits different experimental conditions, sample preparations, and analysis parameters to be

compared in terms of their effects on sequence calling. The method also provides verification and editing functions, which are essential to reading sequences, as well as navigation and analysis tools.

Fig. 8 illustrates the main screen and the associated pull down menus for comparative analysis and visualization of multiple experiments. (SEQ ID NO: 8 and SEQ ID NO: 9) The windows shown are from an appropriately programmed Sun Workstation. However, the comparative analysis software may also be implemented on or ported to a personal computer, including IBM PCs and compatibles, or other workstation environments. A window 802 is shown having pull down menus for the following functions: File 804, Edit 806, View 808, Highlight 810, and Help 812.

The main section of the window is divided into a reference sequence area 814 and a sample sequence area 816. The reference sequence area is where known sequences are displayed and is divided into a reference name subarea 818 and reference base subarea 820. The reference name subarea is shown with the filenames that contain the reference sequences. The chip wild-type is identified by the filename with the extension ".wt#" where the # indicates a unit on the chip. The reference base subarea contains the bases of the reference sequences. A capital C 822 is displayed to the right of the reference sequence that is the chip wild-type for the current analysis. Although the chip wild-type sequence has associated fluorescence intensities, the other reference sequences shown below the chip wild-type may be known sequences that have not been tiled on the chip. These may or may not have associated fluorescence intensities. The reference sequences other than the chip wild-type are used for sequence comparisons and may be in the form of simple ASCII text files.

Sample sequence area 816 is where sample or unknown experimental sequences are displayed for comparison with the reference sequences. The sample sequence area is divided into a sample name subarea 824 and sample base subarea 826. The sample name subarea is shown with filenames that contain the sample sequences. The filename extensions indicate the method used to call the sample sequence where ".cq#" denotes the

intensity ratio method, ".rq#" denotes the reference method, and ".sq#" denotes the statistical method (# indicates the unit on the chip). The sample base subarea contains the bases of the sample sequences. The bases of the sample sequences are identified by the codes previously set forth which, for the most part, conform to the IUPAC standard.

Window 802 also contains a message panel 828. When the user selects a base with an input device in the reference or sample base subarea, the base becomes highlighted and the pathname of the file containing the base is displayed in the message panel. The base's position in the nucleic acid sequence is also displayed in the message panel.

In pull down menu File 804, the user is able to load files of experimental sequences that have been tiled and scanned on a chip. There is a chip wild-type associated with each experimental sequence. The chip wild-type associated with the first experimental sequence loaded is read and shown as the chip wild-type in reference sequence area 814. The user is also able to load files of known nucleic acid sequences as reference sequences for comparison purposes. As before, these known reference sequences may or may not have associated probe intensity data. Additionally, in this menu the user is able to save sequences that are selected on the screen into a project file that can be loaded in at a later time. The project file also contains any linkage of the sequences, where sequences are linked for comparison purposes, ~~individual sequences,~~ sequences to be saved both reference and sample, are ~~selected~~ chosen by selecting the sequence filename with an input device in the reference or sample name subareas.

In pull down menu Edit 806, the user is able to link together sequences in the reference and sample sequence areas. After the user has selected one reference and one or more sample sequences, the sample sequences can be linked to the reference sequence by selecting an entry in the pull down menu. Once the sequences are linked, a link number 830 is displayed next to each of the linked sequences. Each group of linked sequences is associated with a unique link number, so the user can easily identify which sequences are linked

together. Linking sequences permits the user to more easily compare ~~the linked sequences~~ sequences of related interest. The user is also able to remove and display links in this menu.

In pull down menu View 808, the user is able to display intensity graphs for selected bases. Once a base is selected in the reference or sample base subareas, the user may request an intensity graph showing the hybridized probe intensities of the selected base and a delineated neighborhood of bases near the selected base. Intensity graphs may be displayed for one or multiple selected bases. The user is also able to prepare comment files and reports from in this menu.

Fig. 9 illustrates an intensity graph window for a selected base at position 120 (SEQ ID NO:30 and SEQ ID NO:31). The filename containing the sequence data is displayed at 904. The graph shows the intensities for each of the hybridized probes associated with a base. Each grouping of four vertical bars on the graph, which are labeled as "a", "c", "g", and "t" on line 906, shows the background subtracted intensities of probes having the indicated substitution base. In one embodiment, the called bases are shown in red. The wild-type base is shown at line 908, the called base is shown at line 910, and the base position is shown at line 912. In Fig. 9, the base selected is at position 120, as shown by arrow 914. The wild-type base at this position is T; however, the called base is M which means the base is either A or C (amino). The user is able to use intensity graphs to visually compare the intensities of each of the possible calls.

Fig. 10 illustrates multiple intensity graph windows for selected bases (SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34 and SEQ ID NO:35). There are three intensity graph windows 1002, 1004, and 1006 as shown. Each window may be associated with a different experiment, where the sequence analyzed in the experiment may be either a reference (if it has associated probe intensity data as in the chip wild-type) or a sample sequence. The windows are aligned and a rectangular box 1008 shows the selected bases' position in each of the sequences (position 162 in Fig. 10). The rectangular box aids the user in identifying the selected bases.

Referring again to Fig. 8, in pull down menu Highlight 810, the user is able to compare the sequences of references and samples. At least four comparisons are available to the user, including the following: sample sequences to the chip wild-type sequence, sample sequences to any reference sequences, sample sequences to any linked reference sequences, and reference sequences to the chip wild-type sequence. For example, after the user has linked a reference and sample sequence, the user can compare the bases in the linked sequences. Bases in the sample sequence that are different from the reference sequence will then be indicated on the display device to the user (e.g., base is shown in a different color). In another example, the user is able to perform a comparison that will help identify sample sequences. After a sample is linked to multiple reference sequences, each base in the sample sequence that does not match the wild-type sequence is checked to see if it matches one of the linked reference sequences. The bases that match a linked reference sequence will then be indicated on the display device to the user. The user may then more easily identify the sample sequence as being one of the reference sequences.

In pull down menu Help 812, the user is able to get information and instructions regarding the comparative analysis program, the calling methods, and the IUPAC definitions used in the program.

Fig. 11 illustrates the intensity ratio method correctly calling a mutation in solutions with varying concentrations. A window 1102 is shown with a chip wild-type 1104 and a mutant sequence 1106. The mutant sequence differs from the chip wild-type at the position indicated by the rectangular box 1108. The chip wild-type and mutant sequences are a region of HIV Pol Gene spanning mutations occurring in AZT drug therapy.

There are seven sample sequences that are called using the intensity ratio method. The sample sequences are actually solutions of different proportions of the chip wild-type sequence and the mutant sequence. Thus, there are sample

solutions 1110, 1112, 1114, 1116, 1118, 1120, and 1122. The solutions are 15-mer tilings across the chip wild-type with increased percentages of the mutant sequence from 0 to 100% by weight. The following shows the proportions of the sample solutions:

Sample Solution	Chip Wild-Type:Mutant
1110	100:0
1112	90:10
1114	75:25
1116	50:50
1118	25:75
1120	10:90
1122	0:100

For example, sample solution 1114 contains 75% chip wild-type sequence and 25% mutant sequence.

Now referring to the bases called in rectangular box 1108 for the sample solutions, the intensity ratio method correctly calls sample solution 1110 as having a base A as in the chip-wild type sequence. This is correct because sample solution 1110 is 100% chip wild-type sequence. The intensity ratio method also calls sample solution 1112 as having a base A because the sample solution is 90% chip wild-type sequence.

The intensity ratio method calls the identified base in sample solutions 1114 and 1116 as being an R, which is an ambiguity IUPAC code denoting A or G (purine). This also a correct base call because the sample solutions have from 75% to 50% chip-wild type sequence and from 25% to 50% mutation sequence. Thus, the intensity ratio method correctly calls the base in this transition state.

Sample solutions 1118, 1120, and 1122 are called by the intensity ratio method as having a mutation base G at the specified location. This is a correct base call because the sample solutions primarily consist of the mutation sequence (75%, 90%, and 100% respectively). Again, the intensity ratio method correctly called the bases.

These experiments also show that the base calling methods of the present invention may also be used for solutions of more than one nucleic acid sequence.

Fig. 12 illustrates the reference method correctly calling a mutant base where the intensity ratio method incorrectly called the mutant base. There are three intensity (SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, and SEQ ID NO:39)

graph windows 1202, 1204, and 1206 as shown. The windows are aligned and a rectangular box 1208 outlines the bases of interest. Window 1202 shows a sample sequence called using the intensity ratio method. However, the base in the rectangular box 1208 was incorrectly called base ~~C because~~ C, as there is actually a base A at that position. The intensity ratio method incorrectly called the base as C because the probe intensity associated with base C is much higher than the other probe intensities.

Window 1204 shows a reference sequence called using the intensity ratio method. As the reference sequence is known, it is not necessary to know the method used to call the reference sequence. However, it is important to have probe intensities for a reference sequence to use the reference method. The reference sequence ~~has a~~ is called base C at the position indicated by the rectangular box.

Window 1206 shows the sample sequence called using the reference method. The reference method correctly calls the specified base as being base A. Thus, for some cases the reference method is preferable to the intensity ratio method because it compares probe intensities of a sample sequence to probe intensities of a reference sequence.

VII. Examples

Example 1

The intensity ratio method was used in sequence analysis of various polymorphic HIV-1 clones using a protease chip. Single stranded DNA of a 382 nt region was used with 4 different clones (HXB2, SF2, NY5, pPol4mut18). Results were compared to results from an ABI sequencer. The results are illustrated below:

	ABI		Protease Chip	
	Sense	Antisense	Sense	Antisense
No call	0	4	9	4
Ambiguous	6	14	17	8
Wrong call	2	3	3	1
TOTAL	8	21	29	13

SUMMARY
ABI (sense) - 99.5%
Chip (sense) - 98.1%

ABI (antisense) - 98.6%
Chip (antisense) - 99.1%

Example 2

HIV protease genotyping was performed using the described chips and ~~CallSeq~~ CALLSEQ™ intensity ratio calculations. Samples were evaluated from AIDS patients before and after ddI treatment. Results were confirmed with ABI sequencing.

Fig. 13 illustrates the output of the ViewSeq™ program with four pretreatment samples and four posttreatment samples. Note the ~~mutation~~ base change at position 207 where a mutation has arisen. Even adjacent two additional mutations (gt), the "a" mutation has been properly detected.

VIII. Appendices

The Microfiche appendices (copyright Affymetrix, Inc.) provide C++ source code and header files for implementing the present invention. Appendix A contains the source code files (.cc files) for ~~CallSeq~~ CALLSEQ™, which is a base calling program that implements the intensity ratio, reference, and statistical methods of the present invention. Appendix B contains the header files (.h files) for ~~CallSeq~~ CALLSEQ™. Appendices C and D contain the source code and header files, respectively, for a program that performs a preprocessing stage for the statistical method of ~~CallSeq~~ CALLSEQ™.

Appendix E contains the source code and header files for ~~ViewSeq~~ VIEWSEQ™, which is a comparative analysis and visualization program according to the present invention. Appendices A-E are written for a Sun Workstation.

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. Merely by way of example, while the invention is illustrated with particular reference to the evaluation of DNA (natural or unnatural), the methods can be used in the analysis from chips with other materials synthesized thereon, such as RNA. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

85

WHAT IS CLAIMED IS:

CallSeq™

1. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a plurality of probe intensities, each of said probe intensities being associated with a probe on a chip;

said computer system comparing said plurality of probe intensities wherein each of said plurality of probe intensities is substantially proportional to a probe hybridizing with at least one sequence; and

calling said unknown base according to said comparison of said plurality of probe intensities.

2. The method of claim 1, wherein said at least one sequence includes said sample sequence.

3. The method of claim 2, further comprising the step of said computer system calculating a ratio of a higher probe intensity to a lower probe intensity.

4. The method of claim 3, further comprising the step of calling said unknown base as being a base complement of said probe associated with said higher probe intensity if said ratio is greater than a predetermined ratio value.

5. The method of claim 3, wherein said ratio is approximately 1.2.

6. The method of claim 2, further comprising the step of sorting said plurality of probe intensities.

7. The method of claim 1, wherein said at least one sequence includes said sample sequence and a reference sequence.

8. The method of claim 7, further comprising the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to probe intensities hybridizing with said reference sequence.

9. The method of claim 7, further comprising the step of calculating first ratios of a wild-type probe intensity to each probe intensity of a probe hybridizing with said reference sequence, wherein said wild-type probe intensity is associated with a wild-type probe.

10. The method of claim 9, further comprising the step of calculating second ratios of the highest probe intensity of a probe hybridizing with said sample sequence to each probe intensity of a probe hybridizing with said sample sequence.

11. The method of claim 10, further comprising the step of calculating third ratios of said first ratios to said second ratios.

12. The method of claim 7, further comprising the step of comparing neighboring probe intensities of said plurality of probe intensities.

13. The method of claim 7, wherein probe intensities of a probe hybridizing with said reference sequence are from a plurality of experiments.

14. The method of claim 13, further comprising the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to statistics about said plurality of experiments.

15. The method of claim 14, wherein said statistics include a mean and standard deviation.

87

16. The method of claim 13, further comprising the step of normalizing said plurality of probe intensities by dividing each probe intensity by a sum of related probe intensities.

17. The method of claim 1, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

18. The method of claim 1, further comprising the step of setting a probe intensity equal to a relative small positive number if said probe intensity is less than or equal to zero.

19. The method of claim 1, further comprising the step of indicating said unknown base is unable to be called if said plurality of probe intensities have insufficient intensity to call said unknown base.

20. The method of claim 1, wherein said unknown base is called as being A, C, G, or T.

Pooling Processing

21. A method of processing first and second nucleic acid sequences, comprising the steps of:

providing a plurality of nucleic acid probes;

labeling said first nucleic acid sequence with a first marker;

labeling said second nucleic acid sequence with a second marker; and

hybridizing said first and second labeled nucleic acid sequences at the same time.

22. The method of claim 21, wherein said plurality of nucleic acid probes are on a chip.

23. The method of claim 21, further comprising the step of fragmenting said first and second nucleic acid sequences at the same time.

24. The method of claim 21, further comprising the step of scanning for said first and second markers on said chip, said first and second labeled nucleic acid sequences being on said chip.

25. The method of claim 21, wherein said first and second markers are fluorescent markers.

26. The method of claim 25, wherein said first and second markers emit light at different wavelengths upon excitation.

ViewSeq™

27. In a computer system, a method of analyzing a plurality of sequences of bases, said plurality of sequences including at least one reference sequence and at least one sample sequence, the method comprising the steps of:

displaying said at least one reference sequence in a first area on a display device; and

displaying said at least one sample sequence in a second area on said display device;

whereby a user is capable of visually comparing said plurality of sequences.

28. The method of claim 27, wherein said plurality of sequences are monomer strands of DNA or RNA.

29. The method of claim 27, wherein said bases are A, C, G, or T.

30. The method of claim 27, wherein said at least one reference sequence includes a chip wild-type that has been tiled on a chip.

31. The method of claim 30, wherein said chip wild-type sequence is displayed as a first sequence in said first area.

32. The method of claim 30, further comprising the step of displaying a label in said first area to identify said chip wild-type sequence.

33. The method of claim 32, wherein said label is a capital C.

34. The method of claim 27, wherein said at least one sample sequence has been hybridized on a chip.

35. The method of claim 27, further comprising the step of indicating bases that differ among a plurality of user selected sequences.

36. The method of claim 27, further comprising the steps of:

displaying a name associated with each of said at least one reference sequence in said first area; and

displaying a name associated with each of said at least one sample sequence in said second area.

37. The method of claim 27, further comprising the step of linking at least one reference sequence in said first area with at least one sample sequence in said second area.

38. The method of claim 37, further comprising the step of indicating on said display device which sequences are linked.

39. The method of claim 38, wherein said indicating step includes the step of displaying a common symbol next to said linked sequences.

90
50

40. The method of claim 39, wherein said common symbol is a link number.

41. The method of claim 37, further comprising the step of indicating bases of said at least one sample sequence that are not equal to a corresponding base in said at least one reference sequence.

42. The method of claim 27, wherein said at least one reference sequence and said at least one sample sequence are aligned on said display device.

43. The method of claim 27, further comprising the step of exposing sequences to probes.

44. The method of claim 43, further comprising the step of evaluating said exposed sequences according to hybridization with said probes.

add B7

add B7

91
51

COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION

ABSTRACT OF THE DISCLOSURE

A computer system for analyzing nucleic acid sequences is provided. The computer system is used to perform multiple methods for determining unknown bases by analyzing the fluorescence intensities of hybridized nucleic acid probes. The results of individual experiments are improved by processing nucleic acid sequences together. Comparative analysis of multiple experiments is also provided by displaying reference sequences in one area and sample sequences in another area on a display device.

I hereby certify that correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner of Patents and Trademarks, Washington, D.C. 20231, on March 21, 1995

By Anne Currier Carr
Anne Currier Carr

MAR 1995

PATENT

Attorney Docket No. 16528X-82

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
MARK S. CHEE et al.	Examiner: unassigned
Serial No.: 08/327,525	Art Unit: unassigned
Filed: October 21, 1994	PRELIMINARY AMENDMENT
For: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION	

Commissioner of Patents and Trademarks
Washington, D.C. 20231

Sir:

Prior to the examination of the subject application, please amend the application as follows.

IN THE SPECIFICATION:

Please amend the specification as follows.

Page 7, line 8, change "location" to --locations--.

Page 8, line 34, change "tiling. Of course" to --tiling, as described below. Of course,--.

Page 8, line 35, change "herein" to --herein,--.

Page 10, lines 4-5, change "provides the second "chip" above which treats the suspected mutation as the" to --generates the second "chip" above to handle the suspected mutation as a--.

Page 10, lines 12-13, change "control, among other things," to --control--.

Page 11, line 3, change "location" to --position--.

Page 11, line 4, change "other" to --otherwise--.

Page 13, line 1, delete the word "optimally".

Page 13, line 8, after "to", insert --as--.

Page 13, line 12, after "5'-ATGTGGACAGTTGTA-3'", insert --(SEQ ID NO:1)--.

Page 14, line 17, change "by" to --in descending order of--.

Page 14, line 27, change "cutoff" to --cutoff,--.

Page 14, line 34, after "5'-ATGTGGATAGTTGTA-3'", insert --(SEQ ID NO:2)--.

Page 15, line 14, after "ATGTGGAKAGTTGTA-3'", insert --(SEQ ID NO:3)--.

Page 16, lines 12, 18, and 25, change "intensity of base" to --intensities of bases--.

Page 17, line 32, change "wild-type" to --wild-type,--.

Page 18, line 16, after "shown", insert --by the use of two probe lengths--.

Page 18, line 21, after "to", insert --as--.

Page 18, line 25, after "5'-AAAACTGAAAA-3'", insert --(SEQ ID NO:4)--.

Page 18, line 26, after "5'-AAAACCGAAAA-3'", insert --(SEQ ID NO:5)--.

Page 18, line 32, after "5'-AAAACTGAAAA-3'", insert --(SEQ ID NO:4)--.

Page 20, line 1, change "the bases" to --each base--.

Page 23, lines 12, 28, 30, and 37, change "is" to --would be--.

Page 24, line 28, after "5'-AAACCCAATCCACATCA-3'", insert --(SEQ ID NO:6)--.

Page 24, line 29, after "5'-AAACCCAGTCCACATCA-3'", insert --(SEQ ID NO:7)--.

Page 24, line 37, after "Fig. 5B", insert --(SEQ ID NO:6, SEQ ID NO:28, and SEQ ID NO:29)--.

Page 25, line 11, change "umber" to --number--.

Page 26, line 19, change "between" to --across--.

Page 26, line 36, change "ratio" to --ratios--.

Page 26, line 38, change "is" to --are--.

Page 27, lines 5 and 7, change "intensities" to --intensity ratios--.

Page 27, line 30, change "not" to --neither--.

Page 27, line 31, change "and" to --nor--.

Page 28, line 29, change "inot" to --into--.

Page 28, line 30, change "Thus, the" to --The--.

Page 29, lines 10-11, change "good at calling bases if" to --preferable when--.

Page 29, line 12, change "among" to --across--.

Page 29, line 36, after "AAAACTGAAAA-3'", insert --(SEQ ID NO:4)--.

Page 29, line 37, after "5'-AAAACCGAAAA-3'", insert --(SEQ ID NO:5)--.

Page 30, line 3, after "5'-AAAACTGAAAA-3'", insert --(SEQ ID NO:4)--.

Page 31, line 35, change "data" to --data,--.

Page 32, lines 29-30, change "which results in increased" to --improving--.

Page 33, line 5, change "ratio" to --ratios--.

Page 33, line 7, change "is" to --are--.

Page 33, line 11, change "would be" to --are--.

Page 33, line 38, after "The", insert --localized--.

Page 35, line 21, after "with", insert --different--.

Page 35, line 37, delete the word "that" and after "than", insert --that of--.

Page 37, line 7, after "experiments", insert --(SEQ ID NO:8 and SEQ ID NO:9)--.

Page 37, line 19, after "with", insert --the--.

Page 38, line 27, change "Individual sequences, both reference and sample, are selected" to --Sequences to be saved, both reference and sample, are chosen--.

Page 39, line 2, change "the linked sequences." to --sequences of related interest.--

Page 39, line 11, change "in" to --from--.

Page 39, line 13, after "position 120", insert --(SEQ ID NO:30 and SEQ ID NO:31)--.

Page 39, line 23, change "120" to --120,--.

Page 39, line 29, after "bases", insert --(SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, and SEQ ID NO:35)--.

Page 40, line 29, after "concentrations", insert --(SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:18)--.

a1

Page 41, line 41, after "base", insert --(SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, and SEQ ID NO:39)--.

Page 42, line 5, change "C because" to --C, as--.

Page 42, line 15, change "has a" to --is called--.

Page 43, line 24, after "samples", insert --(SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27)--.

Page 43, line 24, change the first occurrence of "mutation" to --base change--.

Please insert the attached pages numbered 45 through 84 and entitled "SEQUENCE LISTING" following page 44 in the specification and renumber original pages 45-51 sequentially as 85-91.

IN THE CLAIMS:

Please amend claims 18 and 19 as follows.

18. (Amended) The method of claim 1, further comprising the step of setting a probe intensity equal to a relatively [relative] small positive number if said probe intensity is less than or equal to zero.

19. (Amended) The method of claim 1, further comprising the step of indicating said unknown base is unable to be called if said plurality of probe intensities has [have] insufficient intensity to call said unknown base.

REMARKS

In reviewing the application, Applicants found some typographical errors, grammar inconsistencies, and the like. Consequently, Applicants present this preliminary amendment. No new matter has been added by these amendments.

Applicants hereby submit a Sequence Listing and a Sequence Listing in computer readable form. Applicants' attorneys attest that the information in the "Sequence Listing" is identical to that which is in computer readable form, as required under 37 CFR 1.821(f). This submission introduces no

new matter, since the enclosed sequences are identical to the sequences which were submitted in the original patent application.

Respectfully submitted,

Michael J. Ritter
Reg. No. 36,653

TOWNSEND and TOWNSEND KHOURIE and CREW
One Market
Steuart Street Tower, 20th Floor
San Francisco, California 94105
(415) 326-2400

MJR/acc
K:\16528\82-PRE.AMD

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231, on Oct. 17, 1995

By Anne Currier Carr

8/B

PATENT

Attorney Docket No. 16528X-82

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
MARKS S. CHEE et al.	Examiner: Unassigned
Serial No.: 08/327,525	Art Unit: 2613
Filed: October 21, 1994	
For: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION	PRELIMINARY AMENDMENT

RECEIVED
96 NOV -9 PM 4:52
GROUP 2600

RECEIVED
NOV 17 1995
GROUP 1800

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

Before examination of the above-identified application on its merits, please amend the application as follows:

IN THE CLAIMS:

Please cancel claims 2 and 21-44 without prejudice. Please amend claims 1, 3-4, 6, 8-12 and 14 as follows.

1. (Amended) In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a plurality of probe intensities, each of said probe intensities being associated with a nucleic acid probe on a chip;

said computer system comparing said plurality of probe intensities wherein each of said plurality of probe intensities is substantially proportional to said associated [a] probe hybridizing with at least one nucleic acid sequence, said at least one nucleic acid sequence including said sample sequence; and

B1

calling said unknown base according to results of said comparing step [said comparison of said plurality of probe intensities].

B2

3. (Amended) The method of claim 1 [2], wherein said comparing step includes [further comprising] the step of said computer system calculating a ratio of a higher probe intensity to a lower probe intensity.

4. (Amended) The method of claim 3, wherein said calling step includes [further comprising] the step of calling said unknown base as being a base according to [complement of] said probe associated with said higher probe intensity if said ratio is greater than a predetermined ratio value.

B3

6. (Amended) The method of claim 1 [2], further comprising the step of sorting said plurality of probe intensities.

B4

8. (Amended) The method of claim 7, wherein said comparing step includes [further comprising] the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to probe intensities hybridizing with said reference sequence.

9. (Amended) The method of claim 7, wherein said comparing step includes [further comprising] the step of calculating first ratios of a wild-type probe intensity to each probe intensity of a probe hybridizing with said reference sequence, wherein said wild-type probe intensity is associated with a wild-type probe.

10. (Amended) The method of claim 9, wherein said comparing step includes [further comprising] the step of calculating second ratios of the highest probe intensity of a probe hybridizing with said sample sequence to each probe intensity of a probe hybridizing with said sample sequence.

B4

11. (Amended) The method of claim 10, wherein said comparing step includes [further comprising] the step of calculating third ratios of said first ratios to said second ratios.

12. (Amended) The method of claim 46 [7], wherein said comparing step includes [further comprising] the step of comparing said ratio of neighboring nucleic acid probes [neighboring probe intensities of said plurality of probe intensities].

B5

14. (Amended) The method of claim 13, wherein said comparing step includes [further comprising] the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to statistics about said plurality of experiments.

Please add claims 45-59 as follows.

B6

-- 45. The method of claim 11, wherein said calling step includes the step of calling said unknown base according to said probe associated with a highest third ratio.

46. The method of claim 7, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity of a probe hybridizing with said reference sequence to a highest intensity of a probe hybridizing with said sample sequence.

47. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a plurality of probe intensities, each of said probe intensities being associated with a nucleic acid probe;

said computer system comparing said plurality of probe intensities wherein each of said plurality of probe intensities

is substantially proportional to said associated nucleic acid probe hybridizing with said sample sequence; and

calling said unknown base according to results of said comparing step.

48. The method of claim 47, wherein said comparing step includes the step of said computer system calculating a ratio of a higher probe intensity to a lower probe intensity.

49. The method of claim 48, wherein said calling step includes the step of calling said unknown base according to said probe associated with said higher probe intensity if said ratio is greater than a predetermined ratio value.

50. The method of claim 49, wherein said predetermined ratio value is approximately 1.2.

51. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a first set of probe intensities, each of said probe intensities in said first set being associated with a nucleic acid probe and substantially proportional to said associated nucleic acid probe hybridizing with a reference nucleic acid sequence;

inputting a second set of probe intensities, each of said probe intensities in said second set being associated with a nucleic acid probe and substantially proportional to said associated nucleic acid probe hybridizing with said sample sequence;

said computer system comparing at least one of said probe intensities in said first set and at least one of said probe intensities in said second set; and

calling said unknown base according to results of said comparing step.

52. The method of claim 51, wherein said comparing step includes the steps of:

calculating first ratios of a wild-type probe intensity to each probe intensity of a probe hybridizing with said reference sequence, wherein said wild-type probe intensity is associated with a wild-type probe; and

calculating second ratios of the highest probe intensity of a probe hybridizing with said sample sequence to each probe intensity of a probe hybridizing with said sample sequence.

53. The method of claim 52, wherein said comparing step further includes the step of calculating third ratios of said first ratios to said second ratios.

54. The method of claim 53, wherein said calling step includes the step of calling said unknown base according to said probe associated with a highest third ratio.

55. The method of claim 51, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity in said first set to a highest intensity in said second set.

56. The method of claim 55, wherein said comparing step further includes the step of comparing said ratio of neighboring nucleic acid probes.

57. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting statistics about a plurality of experiments, each of said experiments producing probe intensities each being associated with a nucleic acid probe and substantially proportional to said associated nucleic acid probe hybridizing with a reference nucleic acid sequence;

inputting a plurality of probe intensities, each of said plurality of probe intensities being associated with a nucleic acid probe and substantially proportional to said associated nucleic acid probe hybridizing with said sample sequence;

said computer system comparing at least one of said plurality of probe intensities with said statistics; and

calling said unknown base according to results of said comparing step.

58. The method of claim 57, further comprising the step of calculating said statistics.

59. The method of claim 57, wherein said statistics include a mean and standard deviation.---

REMARKS

Claims 1, 3-20 and 45-59 are pending in the subject application. Applicants canceled claims 2 and 21-44 without prejudice and reserve all right to pursue these or other claims in another application.

Respectfully submitted,

Michael J. Ritter

Michael J. Ritter
Reg. No. 36,653

TOWNSEND and TOWNSEND and CREW
One Market Plaza
Steuart Street Tower, 20th Floor
San Francisco, California 94105
Tel: (415) 326-2400
Fax: (415) 326-2422

MJR/acc
f:\data\pa\amd\mjr\18661

Dec 04 04 09:22a Kitty Huffman 540-786-0687 p.2

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

SERIAL NUMBER	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.
08/327,525	10/21/94	CHEE	M 165/285S2

EXAMINER
REES, D

ART UNIT	PAPER NUMBER
1807	9

DATE MAILED: 12/19/95

18N1/1219

VERN NORVIEL
TOWNSEND AND TOWNSEND KHOURIE AND CREW
STEUART STREET TOWER
ONE MARKET PLAZA
SAN FRANCISCO CA 94105

This is a communication from the examiner in charge of your application.
COMMISSIONER OF PATENTS AND TRADEMARKS

10-23-95
5/18/95
3/23/95

☑ This application has been examined ☑ Responsive to communication filed on 5/24/95 ☐ This action is made final.

A shortened statutory period for response to this action is set to expire 3 month(s), 0 days from the date of this letter.
Failure to respond within the period for response will cause the application to become abandoned. 35 U.S.C. 133

Part I THE FOLLOWING ATTACHMENT(S) ARE PART OF THIS ACTION:

1. ☑ Notice of References Cited by Examiner, PTO-892.
2. ☐ Notice of Draftsman's Patent Drawing Review, PTO-948.
3. ☑ Notice of Art Cited by Applicant, PTO-1449.
4. ☐ Notice of Informal Patent Application, PTO-152.
5. ☐ Information on How to Effect Drawing Changes, PTO-1474.
6. ☐ _____

Part II SUMMARY OF ACTION

1. ☑ Claims 1, 3-20, 45-59 are pending in the application.
   Of the above, claims _____ are withdrawn from consideration.
2. ☐ Claims _____ have been cancelled.
3. ☐ Claims _____ are allowed.
4. ☑ Claims 1, 3-20, 45-59 are rejected.
5. ☐ Claims _____ are objected to.
6. ☐ Claims _____ are subject to restriction or election requirement.
7. ☐ This application has been filed with informal drawings under 37 C.F.R. 1.85 which are acceptable for examination purposes.
8. ☐ Formal drawings are required in response to this Office action.
9. ☐ The corrected or substitute drawings have been received on _____. Under 37 C.F.R. 1.84 these drawings are ☐ acceptable; ☐ not acceptable (see explanation or Notice of Draftsman's Patent Drawing Review, PTO-948).
10. ☐ The proposed additional or substitute sheet(s) of drawings, filed on _____, has (have) been ☐ approved by the examiner; ☐ disapproved by the examiner (see explanation).
11. ☐ The proposed drawing correction, filed _____, has been ☐ approved; ☐ disapproved (see explanation).
12. ☐ Acknowledgement is made of the claim for priority under 35 U.S.C. 119. The certified copy has ☐ been received ☐ not been received ☐ been filed in parent application, serial no. _____; filed on _____.
13. ☐ Since this application appears to be in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under Ex parte Quayle, 1935 C.D. 11; 453 O.G. 213.
14. ☐ Other

EXAMINER'S ACTION

PTOL-326 (Rev. 2/93)

08327525

Part III DETAILED ACTION

Claim Rejections - 35 USC § 112

1. Claims 1,3-20, 45-59 are rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The following phrases render the claims vague and indefinite:

a) Claim 1 is indefinite in reciting "probe intensities being associated with a nucleic acid probe on a chip" in that it is not clear how the probe intensity is "associated" with the probe. For example, it is not clear whether an "intensity" value is an intrinsic property of each individual probe or if the "intensity" is actually a reflection of the extent of hybridization of probe molecules at a specific site on the chip. The claim might be amended to clarify this point.

b) Claim 1 is further indefinite in reciting "substantially proportional" in that it is not clear how "substantially" is defined.

c) Claim 1 is further indefinite is indefinite in reciting "said associated probe" in that it is the probe intensities which are said to "associated" with a nucleic acid probe; therefore this term lacks proper antecedent basis.

d) Claim 1 is further indefinite in reciting "calling said unknown base" in that it is unclear how "calling" is defined; i.e there appears to be a step missing. The comparing step provides a value with reference to standards? a blank? and the calling step is based on this value? Some recitation should be made in the claims of this intermediate step(s) as it appears to be an essential link between "comparing" and "calling".

e) Claim 4 is indefinite in reciting "calling said unknown base as being a base" in that it is unclear what the unknown base is specifically being "called". The claim might be amended to recite --calling said unknown base an A, T, C or G-- or alternatively --identifying an unknown base--. Further "said probe" lacks proper antecedent basis. It is additionally unclear what "a predetermined ratio value is" in that it is unclear what the reference point for this value is.

f) Claim 6 is indefinite in reciting the "step of sorting" said plurality of probe intensities in that it is not clear what the probe intensities are sorted into (i.e how does this differ from a comparison or the calling step?).

g) Claim 9 is indefinite in reciting "a wild type probe intensity" in that it is not clear how "wild-type" is defined in comparison with the "reference sequence" . Further the recitation of "each probe intensity of a probe" because it is unclear how a single probe can have more than one intensity (as implied by the use of the term "each").

h) Claims 9 and 10 are further indefinite in reciting "first ratios" and "second ratios" in that these ratios are not clearly defined with respect to probes and probe intensities. The problem seems to be mainly one of antecedent basis -it is not clear how "a probe" is to be distinguished from "each probe" in claim 10.

i) Claim 12 is indefinite in reciting "comparing said ratio of neighboring nucleic acid probes" in that it is not clear if the ratios of neighboring nucleic acid probes are compared to each other or to the reference sequence or both.

j) Claim 13 and 14 are indefinite in reciting "Probe intensities of a probe" in that it is not clear how "a" probe generates more than one intensity. It is further how probe intensities are "compared" to statistics and further what the outcome of this step is.

k) Claim 16 is indefinite in reciting "related probe intensities" in that it is not clear how the probes are related.

l) Claim 17 is indefinite in reciting "subtracting a background intensity" in that it is not clear how a background intensity is determined (before hybridization of the probes?).

m) Claim 45 is indefinite in reciting "the step of calling the unknown base". See paragraph 1d.

n) Claim 47 is indefinite in reciting "substantially proportional". See paragraph 1b.

o) Claim 49 is indefinite in reciting "calling step" and "predetermined ratio value". See paragraph 1d and 1e.

p) Claim 51 is indefinite in reciting "substantially proportional" (see paragraph 1b). The claim is further indefinite in reciting "to said associated nucleic acid probe hybridizing with a reference nucleic acid sequence" in that an associated nucleic acid probe hybridizing with a reference sequence has not been previously recited. Similarly the recitation of "said associated nucleic acid probe hybridizing with said sample sequence" lacks proper antecedent basis in this claim. Claim 51 is further indefinite in reciting "calling said unknown base according to results of said comparing step" (paragraph 1d).

q) Claim 52 is indefinite in reciting "calculating first ratios of a wild type probe intensity associated with a wild type probe" , in that it is not clear how a" wild-type probe is defined" (how is it distinguished from the reference sequence?). It is further unclear how "a probe" is distinguished from "each probe".

r) Claim 54 is indefinite in reciting "calling said unknown base". See paragraph 1d.

s) Claim 56 is further indefinite in reciting "comparing the ratio of neighboring nucleic acid probes" in that it is not clear what is being compared: the intensity ratios? And if the latter - of neighboring probes to each other? to reference probes?

t) Claim 57 is unclear in reciting "said plurality of probe intensities being associated with a nucleic acid probe in that it is unclear how a "plurality" of intensities are associate "a"

nucleic acid probe". The claim is further unclear in reciting "comparing at least one of said probe intensities with said statistics"-in that it is not clear exactly what is being compared or what the outcome of this comparison is that allows one to "call" an unknown base sequence.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. § 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.
>
> Subject matter developed by another person, which qualifies as prior art only under subsection (f) or (g) of section 102 of this title, shall not preclude patentability under this section where the subject matter and the claimed invention were, at the time the invention was made, owned by the same person or subject to an obligation of assignment to the same person.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. § 103, the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 C.F.R. § 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order

for the examiner to consider the applicability of potential 35 U.S.C. § 102(f) or (g) prior art under 35 U.S.C. § 103.

Claims 1,3-20, 45-59 are rejected under 35 U.S.C. § 103 as being unpatentable over Fodor et al. WO 92/10588 25 June 1992, in view of Weiss et al (USPAT 5470710, filed Oct 22, 1993)and Stockham et al. (USPAT 5273632, Dec 28, 1993).

Fodor et al. WO 92/10588 25 June 1992, teaches an SBH method wherein initial data resulting from a detection system is an array of data indicative of fluorescent intensity versus location of a substrate. Spurious data points are removed in the method to determine an average of data points. In general the data are fitted to a base curve and statistical measures are used to remove spurious data (page 17, lines 26-40). The detection method provides a positional localization of the region where hybridization takes place and upon having collected all the data indicating the subsequences present in the target sequence, this data may be aligned by overlap to reconstruct the entire sequence of the target (pages 35-36). Fodor also teaches that pixel density may be evaluated over a region to determine the locations and actual extent of a positive signal (this may be interpreted as performing a comparison of intensities in order to "call" a site) (page 76). Fodor teaches that although the method is most directly applicable to sequencing, the invention is also

applicable to fingerprinting, mapping and general screening of specific interactions. Thus the method of Fodor clearly suggests the comparison of hybridization of wild type sequences to mutant sequences or to reference sequences. The method of Fodor et al. differs from that of the present invention in that intensity ratios are not compared as a means on determining the identity of a base, rather it is the location of the signal which is called (although as noted above; distinguishing between different ratios of signal intensities is a part of the method of Fodor which allows one to determine a positive signal at any one site).

Weiss et al (USPAT 5470710, filed Oct 22, 1993) teaches a system which converts the signals obtained from a pattern of multiplex reaction products hybridized with fluorescent probes into a string of nucleotides corresponding to the nucleotide sequence (see abstract).The data acquired is interpreted by an algorithm that yields a "called sequence". Weiss teaches that a CCD snapshot of hybridization signals may be obtained and pixel values may be determined and averaged (column 14, lines 55-63). Ratios of signal intensities are determined using this system and statistics used to calculate standard deviations of sample intensities vs background signals. (see example 7).Further, Stockham et al. (USPAT 5273632, Dec 28, 1993) teaches a method of computerized analysis of the visual images of DNA sequence ladders. A digital lane signal is converted by Fourier transformation to a frequency spectrum. When all the lane signals

from a set of four lanes have been deconvolved using the same lifter function, the signals are normalized to each other (column 4, lines 56-70). A deconvolved signal is subjected to a preliminary peak detection step. A group of putative peaks is established by selecting all peaks which exceed a pre-established threshold intensity. The putative peaks include all peaks whose height (intensity) exceeds the value of the threshold function at their position, i.e a comparison of the ratios of intensities at a segment of a lane is used to determine an average peak height. Determination of threshold values is performed for each of the four lanes and a procedure for registering each of the four lanes is used which preferably places the peaks not only in the correct spatial order, but minimizes the variance of spacing. Preferably the alignment procedure uses high speed sorting across the lanes using a four lane interdependent adjustment of peak positions (columns 9-10). The nucleotide sequence is the correspondence between peak order among the different lanes and the lane associated with each peak, a step referred to as base calling. The methods of Weiss and Stockham do not set any limits on the numbers of ratios that can be determined using their computer system. Therefore it would have been prima facie obvious to one of ordinary skill in the art at the time that the invention was made to use the computer algorithms of Stockham and Weiss to interpret the data inputted from the SBH system of Fodor, given that one could "call" a site

based on the intensity of a signal produced by an associated probe at that site and thus assign an identity to that site. It would be further well within the skill of the ordinary artisan, given the conventionality of standards or reference sequences to determine a predetermined ratio of signal intensities in order to assess whether a positive or negative signal is obtained at that site.

(It therefore appears that the issue of patentability in the present application resides in distinguishing over the "comparing" and "calling" processes of the computer algorithms of Weiss and Stockham and reciting this clearly in the claims).

3. The following references are additionally cited as relevant to programs designed to distinguish between ratios of intensities of light:

Rutenberg et al. (USPAT 4965725,Oct 23, 1993) teaches the use of a neural net system which is a commercially available statistical classifier which identifies a location of interest (in this example a cell) by measurement of integrated optical densities which are the sum of pixel grey values for the object corrected for optical errors. Based on data obtained from a

primary classifier, a secondary classifier is used to check specific areas of the specimen that are deemed to require further screening or classification. Such further examination may be effected by reliance on the already obtained digitized image data for the selected areas of a specimen or by taking additional data components (columns 3, lines 56-60, column 4, lines 7-18) Information within the system is stored in the strength of connections known as weights. In an asynchronous fashion, each processing element computes the sum of products of the weight of each input line multiplied by the signal level (usually 0 or 1) on that input line . If the sum of products exceeds a preset activation threshold, the output processing element is set to 1, if less, it is set to zero. Rutenberg also teaches that "a threelayer neural network can always find a representation that will map any input pattern to any desired output pattern".

Bacus (USPAT 4741043, April 26, 1988) teaches the use of a neural net system to determine the staining of DNA in cytological specimens. The system is calibrated for the optical density of an object , and incoming data may be converted to lookup tables in an imaging processing board so that the output shown optical density can be linearly added to proportionally reflect directly, in this instance the amounts of DNA (column 7, lines 4-11).

4. No claims are allowed.

5. Papers related to this application may be submitted to Group 1800 by facsimile transmission via the P.T.O. Fax Center located in Crystal Mall 1. The CM1 Fax Center number is (703) 305-7939. Please note that the faxing of such papers must conform with the notice to Comply published in the Official Gazette, 1096 OG 30 (Nov 15, 1989).

An inquiry regarding this communication should be directed to examiner Dianne Rees, Ph.D., whose telephone number is (703) 308-6565. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, W. Gary Jones, can be reached on (703) 308-1156.

Calls of a general nature may be directed to the Group receptionist who may be reached at (703) 308-0196.

Dianne Rees
Dianne Rees

Dec 14, 1995

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
12/14/95

I hereby certify that this correspondence is being sent by facsimile transmission to: Examiner D. Rees, Ph.D. Fax No.: 1-703-305-7401 Assistant Commissioner for Patents, Washington, D.C. 20231, on May 20, 1996

By Christina A. Bybee
Christina A. Bybee

PATENT

Attorney Docket No. 16528X-008200
(client file no. 1091)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

MARK S. CHEE ET AL.

Application No.: 08/327,525

Filed: October 21, 1994

For: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION

Examiner: D. Rees

Art Unit: 1807

AMENDMENT

RECEIVED
MAY 20 1996
GROUP 1800

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In response to the Office Action mailed December 19, 1995, for which a petition for an extension of time is enclosed, please amend this application as follows.

IN THE CLAIMS:

Please cancel claims ~~1, 3-20~~ and ~~45-59~~ without prejudice. Please add new claims 60-105 as follows.

1-59. --CANCELED--

60. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence, and each nucleic acid probe differing from each other by a single base;

said computer system comparing said plurality of probe intensities; and

identifying said unknown base according to results of said comparing step.

61. The method of claim 60, wherein said comparing step includes the step of said computer system calculating a ratio of a higher probe intensity to a lower probe intensity.

62. The method of claim 61, wherein said identifying step includes the step of identifying said unknown base according to a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value.

63. The method of claim 62, wherein said predetermined ratio value is approximately 1:2.

64. The method of claim 60, further comprising the step of sorting said plurality of probe intensities before said comparing step.

65. The method of claim 60, wherein said at least one sequence includes a reference sequence.

66. The method of claim 65, wherein said comparing step includes the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to said probe hybridizing with said reference sequence.

67. The method of claim 65, wherein said comparing step includes the step of calculating first ratios of a wild-type probe intensity to each probe intensity of probes hybridizing with said reference sequence, wherein said wild-type probe intensity indicates an extent of hybridization of a complementary probe with said reference sequence.

68. The method of claim 67, wherein said comparing step includes the step of calculating second ratios of the highest probe intensity of a probe hybridizing with said sample sequence to each probe intensity of probes hybridizing with said sample sequence.

69. The method of claim 68, wherein said comparing step includes the step of calculating third ratios of said first ratios to said second ratios.

70. The method of claim 69, wherein said identifying step includes the step of identifying said unknown base according to said probe associated with a highest third ratio.

71. The method of claim 68, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity of a probe hybridizing with said reference sequence to a highest intensity of a probe hybridizing with said sample sequence.

72. The method of claim 71, wherein said comparing step includes the step of comparing said ratio to an equivalent ratio of neighboring nucleic acid probes.

73. The method of claim 65, wherein probe intensities of probes hybridizing with said reference sequence are from a plurality of experiments.

74. The method of claim 73, wherein said comparing step includes the step of said computer system comparing probe intensities of probes hybridizing with said sample sequence to statistics about said plurality of experiments.

75. The method of claim 74, wherein said statistics include a mean and standard deviation.

76. The method of claim 73, further comprising the step of normalizing said plurality of probe intensities by dividing each probe intensity by a sum of related probe intensities, wherein related probe intensities are from probes that differ by a single base.

77. The method of claim 60, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

78. The method of claim 60, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

79. The method of claim 60, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

80. The method of claim 60, wherein said unknown base is identified as being A, C, G, or T.

81. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by a single base;

said computer system calculating a ratio of a higher probe intensity to a lower probe intensity; and

identifying said unknown base according to a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value.

82. The method of claim 81, wherein said predetermined ratio value is approximately 1.2.

83. The method of claim 81, further comprising the step of sorting said plurality of probe intensities before said comparing step.

84. The method of claim 81, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

85. The method of claim 81, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

86. The method of claim 81, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

87. The method of claim 81, wherein said unknown base is identified as being A, C, G, or T.

88. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting a first set of probe intensities, each probe intensity in said first set indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by a single base;

inputting a second set of probe intensities, each probe intensity in said second set indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by a single base;

said computer system comparing at least one of said probe intensities in said first set and at least one of said probe intensities in said second set; and

identifying said unknown base according to results of said comparing step.

89. The method of claim 88, wherein said comparing step includes the step of calculating first ratios of a wild-type probe intensity to each probe intensity of probes hybridizing with said reference sequence, wherein said wild-type probe intensity indicates an extent of hybridization of a complementary probe with said reference sequence.

90. The method of claim 89, wherein said comparing step includes the step of calculating second ratios of the highest probe intensity of probes hybridizing with said sample sequence to each probe intensity of a probe hybridizing with said sample sequence.

91. The method of claim 90, wherein said comparing step further includes the step of calculating third ratios of said first ratios to said second ratios.

92. The method of claim 91, wherein said identifying step includes the step of identifying said unknown base according to said probe associated with a highest third ratio.

93. The method of claim 88, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity in said first set to a highest intensity in said second set.

94. The method of claim 93, wherein said comparing step further includes the step of comparing said ratio to an analogous ratio of neighboring nucleic acid probes.

95. The method of claim 88, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

96. The method of claim 88, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

97. The method of claim 88, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

98. The method of claim 88, wherein said unknown base is identified as being A, C, G, or T.

99. In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

inputting statistics about a plurality of experiments, each of said experiments producing probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by a single base;

inputting a plurality of probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by a single base;

said computer system comparing at least one of said plurality of probe intensities with said statistics; and

identifying said unknown base according to results of said comparing step.

100. The method of claim 99, wherein said statistics include a mean and standard deviation.

101. The method of claim 99, further comprising the step of normalizing said plurality of probe intensities by dividing each probe intensity by a sum of related probe intensities, wherein related probe intensities are from probes that differ by a single base.

102. The method of claim 99, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

103. The method of claim 99, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

104. The method of claim 99, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

105. The method of claim 99, wherein said unknown base is identified as being A, C, G, or T.--

REMARKS

Claims 60-105 are pending in the subject application. In a sincere effort to expedite prosecution Applicants canceled claims 1, 3-20 and 45-59. However, Applicants reserve all right to pursue these or other claims in another application. In light of the amendments and following remarks, Applicants believe all claims now pending are in condition for allowance.

Claims 1, 3-20 and 45-59 were rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject invention. Claims 1, 3-20 and 45-59 were rejected under 35 U.S.C. § 103 as being unpatentable over WO 92/10588 by Fodor et al. ("Fodor") in view of U.S. Patent No. 5,470,710, issued November 28, 1995 to Weiss et al. ("Weiss") and U.S. Patent No.

5,273,632, issued December 28, 1993 to Stockham et al. ("Stockham").

Formal Matters

Applicants appreciate the Examiner's time in discussing the subject application in a telephonic interview on May 20, 1996. In the interview, the Examiner stated that the phrase "relative [sic] small" in claim 78 may be indefinite as it is unclear how relatively small is determined. Applicants changed the claim to delete this phrase so the claim recites that the probe intensity will be set to a positive number if the probe intensity is less than or equal to zero. As discussed in the interview, for a number of different reasons, adjusted probe intensities may become negative or zero. Thus, these probe intensities may be set to a positive number (preferably small) to prevent utilizing negative numbers or zero in future calculations (see page 15, lines 23-29). Applicants similarly changed claims 85, 96 and 103 so Applicants believe that these claims are patentably definite.

The Examiner also requested that Applicants discuss U.S. Patent No. 4,965,725, issued October 23, 1990 to Rutenberg et al. ("Rutenberg") and U.S. Patent No. 4,741,043, issued April 26, 1988 to Bacus. Applicants will discuss these references at the end of this Amendment.

In the Office Action, the Examiner rejected claims 1, 3-20 and 45-59 under § 112, second paragraph, and § 103. In order to expedite prosecution, Applicants canceled these claims rendering the rejections moot. Applicants added new claims and the following paragraphs will show how these claims are allowable over the rejections.

Applicants appreciate the Examiner's careful attention to the pending claims. Although claims 1, 3-20 and 45-59 were canceled, the following will briefly describe how the new claims are patentably definite over the § 112 rejections cited in the Office Action. For the Examiner's convenience, Applicants will label the paragraphs according to the labels in the Office Action.

a) In regard to claim 1, the Examiner stated that it is not clear how a probe intensity is associated with a nucleic acid probe. As the Examiner suggested, Applicants amended claim 60 to recite "each probe intensity indicat[es] an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence." Accordingly, the rejection does not apply to the new claims.

b,c) Also in regard to claim 1, the Examiner stated that "substantially" and "associated" were indefinite or lack antecedent basis. As claim 60 does not contain these words, the rejection does not apply to the new claims.

d) In regard to claim 1, the Examiner stated that it is unclear how "calling" is defined. Claim 60 recites instead "identifying said unknown base" as was suggested by the Examiner in paragraph e). The Examiner also stated that there seems to a step missing. Applicants do not believe that any steps are missing in claim 60. Accordingly, the rejection does not apply to the new claims.

e) In regard to claim 4, the Examiner stated that the phrase "calling said unknown base as being a base" is unclear. Claim 60 recites instead "identifying said unknown base" as suggested by the Examiner. Additionally, the Examiner stated that it is unclear what a "predetermined ratio value" is. A predetermined ratio value is typically a constant number like 1.2 (see, e.g., claim 63). In the interview, it is believed that the Examiner tentatively agreed that this phrase is patentably definite.

f) In regard to claim 6, the Examiner stated that the "step of sorting" is unclear. Claim 64 recites that a step of sorting probe intensities is done "before said comparing step" (see, e.g., page 14, lines 17-22). Accordingly, the rejection does not apply to the new claims.

g) In regard to claim 9, the Examiner stated that it is unclear how "wild-type" is defined with respect to the "reference sequence." Claim 67 recites that the wild-type probe intensity indicates the extent of hybridization of a complementary probe with the reference sequence. Since the reference sequence is a

known sequence, the wild-type probe is also known (see, e.g., page 20, lines 1-7). Also, the Examiner stated that "each probe intensity of a probe" is unclear. Claim 67 recites instead "each probe intensity of probes" (emphasis supplied). Accordingly, the rejection does not apply to the new claims.

h) In regard to claims 9 and 10, the Examiner stated that the phrases "first ratios" and "second ratios" are not clearly defined. The Examiner suggested that the problem is similar to that of claim 9. As claim 68 recites that "each probe intensity of probes hybridizing with said sample sequence" (emphasis supplied), the rejection does not apply to the new claims.

i) In regard to claim 12, the Examiner stated that the phrase "comparing said ratio of neighboring nucleic acid probes" is unclear. Claim 72 recites instead "comparing said ratio to an analogous ratio of neighboring nucleic acid probes" (emphasis supplied). In the interview, the Examiner stated that she understood what Applicants are claiming and would consider if there is a clearer way to recite this in the claims. Applicants invite the Examiner to contact the undersigned if it would aid in prosecution of the subject application.

j) In regard to claims 13 and 14, the Examiner stated that it is not clear how "a" probe generates more than one intensity. Claims 73 and 74 instead contain the plural "probes." Additionally, the Examiner queried how probe intensities may be compared to statistics. One method described in the specification is to compare the probe intensities to a mean and standard deviation (see also claim 75). As to what the result of the comparison will be, this may depend on the implementation of the invention and the data. Accordingly, the rejection does not apply to the new claims.

k) In regard to claim 16, the Examiner stated that it is not clear what is meant by "related probe intensities." Claim 76 recites that "related probe intensities are from probes that differ by a single base" (see also page 31, lines 14-38). Accordingly, the rejection does not apply to the new claims.

1) In regard to claim 17, the Examiner stated that it is unclear how a background intensity is determined. Applicants respectfully point out that it is not necessary that a claim specifically recite "how" each step may be performed. In general, this is the purpose of the specification. Nevertheless, the Examiner stated that the background intensity may be measured before hybridization of the probes. Additionally, the background intensity may be measured from "blank" probes (see, e.g., page 8, lines 27-31). Accordingly, the rejection does not apply to the new claims.

m-t) same as above

The above has shown that the § 112, second paragraph, rejections in the Office Action do not apply to the pending claims. Therefore, Applicants believe that the claims are patentably definite under § 112.

The Invention

The present invention provides innovative computer-aided methods for identifying unknown bases in nucleic acids. The methods compare probe intensities that indicate the extent of hybridization of a nucleic acid probe with a sample nucleic acid, where each of the nucleic acid probes differ from each other by a single base. After comparing the probe intensities, an unknown base is identified (typically as A, C, G, or T) according to the results of the comparison. In one embodiment, a ratio is calculated between the highest probe intensity and the next highest probe intensity. If the ratio is greater than a predetermined ratio value (e.g., 1.2), the unknown base is identified according to nucleic acid probe that produced the highest probe intensity.

The Cited Art Distinguished

Claims 1, 3-20 and 45-58 were rejected under 35 U.S.C. § 103 as being unpatentable over Fodor in view of Weiss and Stockham. Fodor describes, among other things, pioneering techniques for sequencing by hybridization. However, the Examiner cited Weiss and Stockham for disclosing the base calling

(identifying) methods of the present invention. For the following reasons, these references do not disclose or suggest the present invention as claimed.

Weiss and Stockham are related to nucleic acid sequencing which utilizes nucleic acid ladders which may be formed by well known techniques such as the Sanger dideoxy method or the Maxam and Gilbert method. More specifically, Weiss describes utilizing an enzyme on identical probes that hybridize with tags in the fragments of the nucleic acid ladder. The enzymes convert a fluorogenic substrate (e.g., BBTP) into a fluorescent product in order to enhance the pattern of hybridization (see, e.g., Fig. 1C).

Stockham, more specifically, describes methods of sharpening signal peaks from electrophoretic migration patterns of nucleic acid ladders. Each fragment of the nucleic acid ladder is labeled with a radioactive label which is utilized to identify the position of the fragment on the gel following electrophoresis. As analyzing the migration patterns is time consuming and often error prone, Stockham describes equations and formulas for increasing the accuracy of this process (e.g., sharpening signal peaks).

Weiss and Stockham do not disclose or suggest inputting probe intensities to identify an unknown base where the probe intensities indicate the extent of hybridization of probes differing by a single base and the sample nucleic acid sequence. Claim 60 recites the following:

> inputting a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence, and each nucleic acid probe differing from each other by a single base;

(emphasis supplied). Neither Weiss nor Stockham discloses these limitations.

Initially, Weiss uses a single probe which will hybridize to a tag on the nucleic acid ladder fragments. As such, all of the "probes" in Weiss are identical. Furthermore, the probes in Weiss do not indicate the extent of hybridization but instead are utilized to generate a fluorescent signal which

indicates the location of a fragment on the substrate. Accordingly, it is the location of the fragments that is utilized to sequence a nucleic acid.

Stockham does not utilize probes at all. Instead, Stockham recites that the fragments of the nucleic acid ladder are radioactively labeled. The radioactive signal resulting indicates the position of the fragments on the gel in a way which is similar to Weiss. Accordingly, Stockham also utilizes the location of the fragments to sequence a nucleic acid.

In stark contrast, the present invention compares probe intensities that indicate the extent of hybridization of probes differing by a single base and the sample nucleic acid sequence. Claim 60 recites the following:

> said computer system comparing said plurality of probe intensities; and
> identifying said unknown base according to results of said comparing step.

In the Office Action, the Examiner stated that it would have been prima facie obvious to one of ordinary skill in the art to use the computer algorithms of Weiss and Stockham to interpret that data from the sequencing by hybridization described by Fodor. More specifically, the Examiner stated that one could "call" a site based on the intensity of a signal produced by a probe at that site and thus assign an identity to that site. Applicants disagree.

Weiss and Stockham relate to vastly different technologies than the pioneering advances of Fodor. Weiss and Stockham are directed to identifying the location of a fragment of a nucleic acid ladder. In the present invention, the locations of the hybridized probes are known and, as such, the computer algorithms of Weiss and Stockham would indeed seem to teach away from the present invention which is directed to calling an unknown base according to probe intensities from nucleic acid probes that differ by a single base.

As Weiss and Stockham do not disclose or suggest all the limitations of claim 60, the claim is patentably distinct over the references. All the other pending claims contain similar limitations. Therefore, Applicants request that all the

pending claims be passed to issue.

Other Claims

Independent claims 81, 88 and 99 recite specific methods of identifying unknown bases. Details on specific embodiments of these methods may be found in the specification under the headings "Intensity Ratio Method," "Reference Method" and "Statistical Method." These claims recite methods that are patentable for at least the same reasons as above. Additionally, these claims include further limitations that make them further patentably distinct.

Claim 81 recites that a ratio of a higher probe intensity and a lower probe intensity is calculated. Then, the unknown base is identified according to the probe that had the higher probe intensity if the ratio is greater than a predetermined ratio value. Weiss and Stockham simply do not disclose or suggest this method. Accordingly, claims 81-87 are patentably distinct.

Claim 88 recites that probe intensities from a first set of probe intensities from probes hybridizing with a reference nucleic acid sequence and a second set of probe intensities from probes hybridizing with a sample nucleic acid sequence are compared. Based on this comparison, the unknown base is identified. Weiss and Stockham do not disclose or suggest this method. Accordingly, claims 88-98 are patentably distinct.

Claim 99 recites that a probe intensity of a nucleic acid probe hybridizing with a sample sequence is compared to statistics from nucleic acid probes hybridizing with a reference sequence. Based on this comparison, the unknown base is identified. Weiss and Stockham do not disclose or suggest this method. Accordingly, claims 99-105 are patentably distinct.

Additionally Cited Art

In the Office Action, the Examiner cited Rutenberg and Bacus as relevant to programs designed to distinguish ratios of intensities of light. Although in the interview the Examiner stated that these references may be nonanalogous art, she

requested that these references be discussed by Applicants. The following will show that these references do not teach or suggest the present invention regardless of whether the references are analogous art.

Rutenberg describes a two stage neural network system for classifying cells on a slide, e.g., for detecting cervical cancer. In a first stage, the neural network classifies cells or objects which are pre-malignant and malignant. However, the first stage may include other nonmalignant objects like cell clumps, debris, leucocytes, and mucus. A second stage of the neural network is utilized to distinguish the pre-malignant and malignant cells from the nonmalignant objects. As Rutenberg describes methods of distinguishing objects on a slide utilizing neural networks, the reference does not disclose or suggest the base calling methods of the present invention.

Bacus describes a method for overcoming staining variations among cells for analysis, e.g., for cancer diagnosis and prognosis. Conventional staining mechanisms may have variations among experiments so reference cells are placed on the slides with the specimen cells. After staining, the imaging apparatus is calibrated according to the reference cells. The specimen cells are then analyzed to determine characteristics such as nuclear optical density. As Bacus describes methods of calibrating imaging apparatus for analyzing cells on a slide, the reference does not disclose or suggest the base calling methods of the present invention.

CONCLUSION

In view of the foregoing, Applicants believe all claims now pending in this application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

If the Examiner believes a telephone conference would expedite prosecution of this application, please telephone the undersigned at (415) 326-2400.

Respectfully submitted,

Michael J. Ritter
Reg. No. 36,653

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 326-2400
Fax (415) 326-2422

MJR/cab

k:\1652\182\008200.am2

Atty Docket No. 16528X-008200

PTO FAX NO.: 1-703-305-7401

ATTENTION: EXAMINER D. REES, PH.D., ART UNIT 1807

CERTIFICATION OF FACSIMILE TRANSMISSION

I hereby certify that the following Amendment, in re Application of Mark S. Chee, Application No. 08/327,525, filed October 21, 1994, for COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION is being transmitted by facsimile to the Patent and Trademark Office on the date shown below.

Number of pages being transmitted, including this page: 23

Dated: May 20, 1996

Christine A. Bybee

Christine A. Bybee

PLEASE CONFIRM RECEIPT OF THIS PAPER BY
RETURN FACSIMILE AT (415) 576-0300

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, CA 94111-3834
Telephone: (415) 326-2400
Fax: (415) 326-2422

M:\16528\8\FAX.COV

MAY 20 '96 04:48PM TTC PALO ALTO 415 326 2422 P.2/24

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, CA 94111-3834
(415) 576-0200

Atty. Docket No. 16528-008200

Date May 20, 1996

In re application of Mark S. Chee et al.

Appln. No. 08/327,525

Filed October 21, 1994

Group Art Unit 1807
For: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION

I hereby certify that this correspondence is being sent by facsimile transmission to: Examiner D. Rees, Ph.D.
Fax No.: 1-703-305-7401
Assistant Commissioner for Patents,
Washington, D.C. 20231
on May 20, 1996
by: Christine A. Bybee
Christine A. Bybee

THE ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231

Sir:

Transmitted herewith is an amendment in the above-identified application.

[X] Enclosed is a petition to extend time to respond.
If any extension of time is needed, then this response should be considered a petition therefor.
The filing fee has been calculated as shown below:

	(Col. 1) CLAIMS REMAINING AFTER AMENDMENT		(Col. 2) HIGHEST NO. PREVIOUSLY PAID FOR	(Col. 3) PRESENT EXTRA	SMALL ENTITY RATE	SMALL ENTITY ADDIT. FEE	OR	OTHER THAN A SMALL ENTITY RATE	OTHER THAN A SMALL ENTITY ADDIT. FEE
TOTAL	*46	MINUS	**44	=2	x11=	$		x22=	$44
INDEP.	*4	MINUS	***3	=1	x39=	$		x78=	$78
[ ] FIRST PRESENTATION OF MULTIPLE DEP. CLAIM					+125=	$		+250=	$
					TOTAL ADDIT. FEE	$	OR	TOTAL	$122

* If the entry in Col. 1 is less than the entry in Col. 2, write "0" in Col. 3.

** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, write "20" in this space.

*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, write "3" in this space.
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found from the equivalent box in Col. 1 of a prior amendment or the number of claims originally filed.

[ ] No fee is due.

Please charge Deposit Account No. 20-1430 as follows:

[X] Claims fee $ 122

[X] Any additional fees associated with this paper or during the pendency of this application.

2 extra copies of this sheet are enclosed.

TOWNSEND and TOWNSEND and CREW LLP

Michael J. Ritter
Reg. No.: 36,653
Attorneys for Applicant

GR 1807

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231, on 1-9-97
By Christine A. Bykel

PATENT

Attorney Docket No. 16528X-008200
(client file no. 1091)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
MARK S. CHEE ET AL.	Examiner: D. Rees
Application No.: 08/327,525	Art Unit: 1807
Filed: October 21, 1994	
For: COMPUTER-AIDED VISUALIZATION AND ANALYSIS SYSTEM FOR SEQUENCE EVALUATION	AMENDMENT

15 D dft
Aug 98
1-9-97

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In response to the Office Action mailed July 9, 1996, for which a petition for an extension of time is enclosed, please amend this application as follows.

IN THE CLAIMS:

For the Examiner's convenience, all claims pending are shown below. Claims that have not been amended herein are shown in small print.

Please cancel claims 72 and 94 without prejudice and amend claims 60, 62, 64, 70, 72, 81, 83, 88, 92, 94, and 99 as follows.

1-59. ~~CANCELED~~ Done by C.

60. (Amended) In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

imputing a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

said computer system comparing said plurality of probe intensities to each other; and

said computer system generating a base call identifying said unknown base according to results of said comparing step.

61. The method of claim 60, wherein said comparing step includes the step of said computer system calculating a ratio of a higher probe intensity to a lower probe intensity.

62. (Amended) The method of claim 61, wherein said generating [identifying] step includes the step of identifying said unknown base according to a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value.

63. The method of claim 62, wherein said predetermined ratio value is approximately 1.2.

64. (Amended) The method of claim 60, further comprising the step of sorting said plurality of probe intensities by intensity before said comparing step.

65. The method of claim 60, wherein said at least one sequence includes a reference sequence.

66. The method of claim 65, wherein said comparing step includes the step of said computer system comparing probe intensities of a probe hybridizing with said sample sequence to said probe hybridizing with said reference sequence.

67. The method of claim 65, wherein said comparing step includes the step of calculating first ratios of a wild-type probe intensity to each probe intensity of probes hybridizing with said reference sequence, wherein

said wild-type probe intensity indicates an extent of hybridization of a complementary probe with said reference sequence.

68. The method of claim 67, wherein said comparing step includes the step of calculating second ratios of the highest probe intensity of a probe hybridizing with said sample sequence to each probe intensity of probes hybridizing with said sample sequence.

69. The method of claim 68, wherein said comparing step includes the step of calculating third ratios of said first ratios to said second ratios.

70. (Amended) The method of claim 69, wherein said generating [identifying] step includes the step of identifying said unknown base according to a base of said probe associated with a highest third ratio.

71. The method of claim 68, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity of a probe hybridizing with said reference sequence to a highest intensity of a probe hybridizing with said sample sequence.

72. --CANCELED--

73. The method of claim 65, wherein probe intensities of probes hybridizing with said reference sequence are from a plurality of experiments.

74. The method of claim 73, wherein said comparing step includes the step of said computer system comparing probe intensities of probes hybridizing with said sample sequence to statistics about said plurality of experiments.

75. The method of claim 74, wherein said statistics include a mean and standard deviation.

76. The method of claim 73, further comprising the step of normalizing said plurality of probe intensities by dividing each probe intensity by a sum of related probe intensities, wherein related probe intensities are from probes that differ by a single base.

77. The method of claim 60, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

78. The method of claim 60, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

79. The method of claim 60, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

80. The method of claim 60, wherein said unknown base is identified as being A, C, G, or T.

81. (Amended) In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

imputing a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

said computer system calculating a ratio of a higher probe intensity to a lower probe intensity; and

said computer system generating a base call identifying said unknown base according to a base of a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value.

82. The method of claim 81, wherein said predetermined ratio value is approximately 1.2.

83. (Amended) The method of claim 81, further comprising the step of sorting said plurality of probe intensities by intensity before said comparing step.

84. The method of claim 81, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

85. The method of claim 81, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

86. The method of claim 81, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

87. The method of claim 81, wherein said unknown base is identified as being A, C, G, or T.

88. (Amended) In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

imputing a first set of probe intensities, each probe intensity in said first set indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

imputing a second set of probe intensities, each probe intensity in said second set indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

said computer system comparing at least one of said probe intensities in said first set and at least one of said probe intensities in said second set; and

said computer system generating a base call identifying said unknown base according to results of said comparing step.

89. The method of claim 88, wherein said comparing step includes the step of calculating first ratios of a wild-type probe intensity to each probe intensity of probes hybridizing with said reference sequence, wherein said wild-type probe intensity indicates an extent of hybridization of a complementary probe with said reference sequence.

90. The method of claim 89, wherein said comparing step includes the step of calculating second ratios of the highest probe intensity of probes hybridizing with said sample sequence to each probe intensity of a probe hybridizing with said sample sequence.

91. The method of claim 90, wherein said comparing step further includes the step of calculating third ratios of said first ratios to said second ratios.

92. (Amended) The method of claim 91, wherein said generating [identifying] step includes the step of identifying said unknown base according to a base of said probe associated with a highest third ratio.

93. The method of claim 88, wherein said comparing step includes the step of calculating a ratio of a highest probe intensity in said first set to a highest intensity in said second set.

94. --CANCELED--

95. The method of claim 88, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

96. The method of claim 88, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

97. The method of claim 88, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

98. The method of claim 88, wherein said unknown base is identified as being A, C, G, or T.

99. (Amended) In a computer system, a method of identifying an unknown base in a sample nucleic acid sequence, said method comprising the steps of:

imputing statistics about a plurality of experiments, each of said experiments producing probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

imputing a plurality of probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid

probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

said computer system comparing at least one of said plurality of probe intensities with said statistics; and

said computer system generating a base call identifying said unknown base according to results of said comparing step.

100. The method of claim 99, wherein said statistics include a mean and standard deviation.

101. The method of claim 99, further comprising the step of normalizing said plurality of probe intensities by dividing each probe intensity by a sum of related probe intensities, wherein related probe intensities are from probes that differ by a single base.

102. The method of claim 99, further comprising the step of subtracting a background intensity from each of said plurality of probe intensities.

103. The method of claim 99, further comprising the step of setting a probe intensity equal to a positive number if said probe intensity is less than or equal to zero.

104. The method of claim 99, further comprising the step of indicating said unknown base is unable to be identified if said plurality of probe intensities have insufficient intensity to identify said unknown base.

105. The method of claim 99, wherein said unknown base is identified as being A, C, G, or T.

Please add new claims 106-117 as follows.

106. The method of claim 60, wherein the plurality of nucleic acid probes are in an array of probes.

107. The method of claim 60, wherein the plurality of probe intensities are fluorescent intensities.

108. A computer program product that identifies an unknown base in a sample nucleic acid sequence, comprising:

computer code that receives a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of said plurality of probe intensities to each other;

computer code that generates a base call identifying said unknown base according to results of said comparison; and

a computer readable medium that stores said computer codes.

109. A computer program product that identifies an unknown base in a sample nucleic acid sequence, comprising:

computer code that receives a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that calculates a ratio of a higher probe intensity to a lower probe intensity;

computer code that generates a base call identifying said unknown base according to a base of a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value; and

a computer readable medium that stores said computer codes.

110. A computer program product that identifies an unknown base in a sample nucleic acid sequence, comprising:

computer code that receives a first set of probe intensities, each probe intensity in said first set indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that receives a second set of probe intensities, each probe intensity in said second set indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of at least one of said probe intensities in said first set and at least one of said probe intensities in said second set;

computer code that generates a base call identifying said unknown base according to results of said comparison; and

a computer readable medium that stores said computer codes.

111. A computer program product that identifies an unknown base in a sample nucleic acid sequence, comprising:

computer code that receives statistics about a plurality of experiments, each of said experiments producing probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that receives a plurality of probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of at least one of said plurality of probe intensities with said statistics;

computer code that generates a base call identifying said unknown base according to results of said comparison; and

a computer readable medium that stores said computer codes.

112. A system that identifies an unknown base in a sample nucleic acid sequence, comprising:

a processor; and

a computer readable medium coupled to said processor for storing a computer program comprising:

computer code that receives a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with at least one nucleic acid sequence including said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of said plurality of probe intensities to each other; and

computer code that generates a base call identifying said unknown base according to results of said comparison.

113. A system that identifies an unknown base in a sample nucleic acid sequence, comprising:

a processor; and

a computer readable medium coupled to said processor for storing a computer program comprising:

computer code that receives a plurality of probe intensities for a plurality of nucleic acid probes, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that calculates a ratio of a higher probe intensity to a lower probe intensity; and

computer code that generates a base call identifying said unknown base according to a base of a nucleic acid probe having said higher probe intensity if said ratio is greater than a predetermined ratio value.

114. A system that identifies an unknown base in a sample nucleic acid sequence, comprising:

a processor; and

a computer readable medium coupled to said processor for storing a computer program comprising:

computer code that receives a first set of probe intensities, each probe intensity in said first set indicating an

extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that receives a second set of probe intensities, each probe intensity in said second set indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of at least one of said probe intensities in said first set and at least one of said probe intensities in said second set; and

computer code that generates a base call identifying said unknown base according to results of said comparison.

115. A system that identifies an unknown base in a sample nucleic acid sequence, comprising:

a processor; and

a computer readable medium coupled to said processor for storing a computer program comprising:

computer code that receives statistics about a plurality of experiments, each of said experiments producing probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with a reference nucleic acid sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that receives a plurality of probe intensities, each probe intensity indicating an extent of hybridization of a nucleic acid probe with said sample sequence, and each nucleic acid probe differing from each other by at least a single base;

computer code that performs a comparison of at least one of said plurality of probe intensities with said statistics; and

computer code that generates a base call identifying said unknown base according to results of said comparison.--

116. A system according to claims 112, 113, 114, or 115, wherein the plurality of nucleic acid probes are in an array of probes.

117. A system according to claims 112, 113, 114, or 115, wherein the plurality of probe intensities are fluorescent intensities.--

REMARKS

Claims 60-71, 73-93, and 95-117 are pending in the subject application. Applicants canceled claims 72 and 94 without prejudice and reserve all right to pursue these or other claims in another application. Claims 106-117 were added by this Amendment. In light of the amendments and following remarks, Applicants believe all claims now pending are in condition for allowance.

The Examiner indicated that claims 60-105 were allowable over the cited art, however, the disclosure was objected to because the Sequence Listing did not conform to 37 C.F.R. §§ 1.821-25. Additionally, claims 60-105 were rejected under 35 U.S.C. § 101 as being directed at non-statutory subject matter and under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject invention.

The Sequence Listing

The Office Action mailed July 9, 1996 did not include a Notice to Comply. Applicants telephoned the Examiner on July 17, 1996 to inform her that the Notice had not been received. Applicants appreciate the Examiner's diligence in faxing the Sequence Verification Report (enclosed) to the undersigned's attention on August 22, 1996.

Applicants corrected the Sequence Listing to remedy errors specified in the faxed Report. The corrections include replacing each occurrence of 'X' in the Sequence Listing with an 'N' (both were utilized in the specification to indicate an

ambiguous base; see page 11, lines 38-42, of the specification). Applicants submit herewith a corrected Sequence Listing and diskette including the Sequence Listing in computer readable form.

The Office Action stated that the Notice to Comply had a shortened statutory period for response of one month from the mailing date; however, Applicants never received the Notice. In a telephone conversation with the Examiner on December 9, 1996, the Examiner indicated that it would be permissible to file the new Sequence Listing with this Amendment along with the requisite petition for three months extension of time. Should any other fees be required, Applicants authorize any other fees to be charged to Deposit Account No. 20-1430.

The Invention

The Examiner acknowledges that the present invention provides innovative computer-aided methods that are not disclosed or suggested in the prior art for identifying unknown bases in nucleic acids. However, the Examiner rejected the claims under § 101 and § 112, second paragraph. Applicants have amended the claims to overcome the rejections and will discuss each of the Examiner's specific rejections below.

The § 101 Rejection

The Examiner rejected claims 60-105 under 35 U.S.C. § 101 as being directed to non-statutory subject matter. More specifically, the Examiner stated that the claims are directed to a computer algorithm which is not applied to physical elements or process steps. Applicants respectfully disagree. The claims recite computer-aided processes which analyze probe intensities indicative of the extent of hybridization of the nucleic acid probes and the sample sequence. The extent of hybridization is indeed a physical quantity just as are electrocardiograph signals (see Arrhythmia Research Tech. v. Corazonix Corp., 958 F2d 1053, 22 USPQ2d 1033 (Fed. Cir. 1992)). In order to more clearly recite the invention, Applicants amended the independent claims to recite that "said computer system generat[es] a base call"

that identifies an unknown base. Therefore, the claims recite statutory computer-aided processes of receiving probe intensities indicative of hybridization with a sample sequence and generating a base call of an unknown base in the sample sequence.

The Examiner suggested that the claims "should be amended so that if viewed without the algorithm the process would stand alone." Applicants understand the Examiner to say that the claims should recite statutory subject matter in the abstract, without consideration of an algorithm or computer. For example, processes of curing synthetic rubber are statutory subject matter so a process of curing synthetic rubber that utilizes a computer is also statutory subject matter (see Diamond v. Diehr, 450 US 175, 209 USPQ 1 (1981)). Applicants submit that processes for identifying unknown bases in sample nucleic acids are statutory subject matter so the recited computer-aided processes are also statutory subject matter.

Applicants have reviewed the Examination Guidelines for Computer-Related Inventions in § 2106 of the MPEP (see, e.g., the section entitled "Manipulation of Data Representing Physical Objects or Activities" in MPEP § 2106(IV)(B)(2)(b)(i)). Applicants fully believe that the pending claims are directed to statutory subject matter. Applicants invite the Examiner to telephone the undersigned if a telephone discussion would facilitate prosecution of the subject application.

The § 112, Second Paragraph, Rejections

The Examiner rejected claims 60-105 under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject invention. With regard to claim 60, the Examiner indicated it is not clear to what the probe intensities are compared. Applicants amended claim 60 to recited that the probe intensities are compared to each other (see, e.g., Figs. 3, 4A and 5A). Accordingly, the rejection is overcome.

Additionally, the Examiner stated that claim 60 is indefinite in that it is not clear how one extrapolates from "comparing" to "identifying." Applicants respectfully submit

that, when read in light of the specification as the case law requires, the claims are not unclear. Applicants' specification provides full detail on possible "comparing" and "generating" (as amended) steps. For example, the highest probe intensity may be compared to the next highest probe intensity to generate a ratio. If this ratio is greater than a predetermined ratio cutoff, the unknown base will be called according to (e.g., complementary to) a base in the probe with the highest intensity (see page 14, lines 23-34). Many of the dependent claims recite further details on these steps (see, e.g., claims 61-63). As Applicants' specification provides ample description of exemplary ways that the present invention may be performed, Applicants request that this rejection be withdrawn.

With regard to claim 64, the Examiner indicated it is unclear how the probe intensities are sorted. Applicants amended claim 64 to recite that the probe intensities are sorted "by intensity" (see, e.g., page 14, line 17). Accordingly, the rejection is overcome.

The Examiner indicated that in claim 70 it is unclear what characteristic of the probe "identifies" the unknown base. As described in the specification, the nucleic acid probes are generally complementary to the sample sequence in order to allow for hybridization between the probes and the sample sequence. Applicants amended claim 70 to recite that the unknown base is identified according to "a base" of a probe with a highest ratio (see page 24, lines 5-10). Accordingly, the rejection is overcome.

The Examiner rejected claims 72 and 94. In a sincere effort to expedite prosecution. Applicants canceled these claims rendering the rejections moot.

The Examiner rejected claims 81, 83, and 92 for the same reasons as claims discussed above. Applicants amended these and other claims in the manner discussed to overcome the rejections. Therefore, Applicants have addressed all the § 112, second paragraph, rejections and respectfully request that these rejections be withdrawn.

CONCLUSION

In view of the foregoing, Applicants believe all claims now pending in this application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

If the Examiner believes a telephone conference would expedite prosecution of this application, please telephone the undersigned at (415) 326-2400.

Respectfully submitted,

Michael J. Ritter
Reg. No. 36,653

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(415) 326-2400
Fax (415) 326-2422

MJR:cab

K:\11652\1\2000Z00.amd

Notice of Allowability	Application No. 08/327,525	Applicant(s) Chee et al.
	Examiner Dianne Rees	Group Art Unit 1807

All claims being allowable, PROSECUTION ON THE MERITS IS (OR REMAINS) CLOSED in this application. If not included herewith (or previously mailed), a Notice of Allowance and Issue Fee Due or other appropriate communication will be mailed in due course.

☒ This communication is responsive to *fax of 4/28/97*.

☒ The allowed claim(s) is/are *108-117*.

☐ The drawings filed on _____ are acceptable.

☐ Acknowledgement is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d).

☐ All ☐ Some* ☐ None of the CERTIFIED copies of the priority documents have been

☐ received.

☐ received in Application No. (Series Code/Serial Number) _____.

☐ received in this national stage application from the International Bureau (PCT Rule 17.2(a)).

*Certified copies not received: _____.

☐ Acknowledgement is made of a claim for domestic priority under 35 U.S.C. § 119(e).

A SHORTENED STATUTORY PERIOD FOR RESPONSE to comply with the requirements noted below is set to EXPIRE THREE MONTHS FROM THE "DATE MAILED" of this Office action. Failure to timely comply will result in ABANDONMENT of this application. Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

☐ Note the attached EXAMINER'S AMENDMENT or NOTICE OF INFORMAL APPLICATION, PTO-152, which discloses that the oath or declaration is deficient. A SUBSTITUTE OATH OR DECLARATION IS REQUIRED.

☒ Applicant MUST submit NEW FORMAL DRAWINGS

☒ because the originally filed drawings were declared by applicant to be informal.

☐ including changes required by the Notice of Draftsperson's Patent Drawing Review, PTO-948, attached hereto or to Paper No. _____.

☐ including changes required by the proposed drawing correction filed on _____, which has been approved by the examiner.

☐ including changes required by the attached Examiner's Amendment/Comment.

Identifying indicia such as the application number (see 37 CFR 1.84(c)) should be written on the reverse side of the drawings. The drawings should be filed as a separate paper with a transmittal letter addressed to the Official Draftsperson.

☐ Note the attached Examiner's comment regarding REQUIREMENT FOR THE DEPOSIT OF BIOLOGICAL MATERIAL.

Any response to this letter should include, in the upper right hand corner, the APPLICATION NUMBER (SERIES CODE/SERIAL NUMBER). If applicant has received a Notice of Allowance and Issue Fee Due, the ISSUE BATCH NUMBER and DATE of the NOTICE OF ALLOWANCE should also be included.

Attachment(s)

☒ Notice of References Cited, PTO-892

☐ Information Disclosure Statement(s), PTO-1449, Paper No(s). _____

☐ Notice of Draftsperson's Patent Drawing Review, PTO-948

☐ Notice of Informal Patent Application, PTO-152

☒ Interview Summary, PTO-413

☒ Examiner's Amendment/Comment

☐ Examiner's Comment Regarding Requirement for Deposit of Biological Material

☐ Examiner's Statement of Reasons for Allowance

F

DETAILED ACTION

1. An examiner's amendment to the record appears below. Should the changes and/or additions be unacceptable to applicant, an amendment may be filed as provided by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview with Vern Norveil and Nancy De Santis on 4/28/97.

2. The application has been amended as follows:

Claims 60-71, 73-93, 95-107 have been canceled.

In claim 110, line 11, "correspodning" has been deleted and --corresponding-- inserted.

In claim 114, line 21, after "said", --sequence of said -- has been inserted.

In claim 118, line 22, after "said", --sequence of said-- has been inserted.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dianne Rees whose telephone number is (703) 308-6565.

W. GARY JONES
SUPERVISORY PATENT EXAMINER
GROUP 1800
4/28/97

Dianne Rees
April 28, 1997